

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
Приладобудівний факультет
Кафедра приладів і систем орієнтації і навігації**

До захисту допущено:

Завідувач кафедри

_____ Надія БУРАУ

«__» _____ 20__ р.

Дипломна робота

на здобуття ступеня бакалавра

**за освітньо-професійною програмою «Комп'ютерно - інтегровані технології
та системи навігації і керування»**

спеціальності 151 «Автоматизація та комп'ютерно-інтегровані технології»

**на тему: «Методичне та алгоритмічне забезпечення для прогнозування
фізичних процесів на основі нейромережових технологій»**

Виконав (-ла):

студент (-ка) IV курсу, групи ПГ-61

Дорошенко Антон Олександрович _____

Керівник:

Зав. каф., д.т.н., проф.

Бурау Надія Іванівна _____

Рецензент:

Доцент, к.т.н.

ГОЛУБЄВ Леонтій Петрович _____

Засвідчую, що у цій дипломній роботі
немає запозичень з праць інших авторів
без відповідних посилань.

Студент (-ка) _____

Київ – 2020 року

**Національний технічний університет України
«Київський політехнічний інститут
імені Ігоря Сікорського»**

Інститут (факультет) _____ Приладобудівний _____
(повна назва)

Кафедра _____ Приладів і систем орієнтації і навігації _____
(повна назва)

Рівень вищої освіти – перший (бакалаврський)

Спеціальність _____ 151 – Автоматизація та комп'ютерно-інтегровані технології _____
(код і назва)

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Н.І. Бурау _____
(підпис) (ініціали, прізвище)

«___» _____ 20__ р.

ЗАВДАННЯ

на дипломну роботу студенту

_____ Дорошенку Антону Олександровичу _____

(прізвище, ім'я, по батькові)

1. Тема дипломної роботи: _____ Методичне та алгоритмічне забезпечення для прогнозування фізичних процесів на основі нейромережових технологій _____

Науковий керівник роботи _____ Бурау Надія Іванівна, д.т.н., професор _____ ,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від «___» _____ 201__ р. № _____

2. Термін подання студентом дисертації «09» червня 2020 р.

3. Вихідні дані до роботи Процеси: 1) монотонні (kt , $(1-\exp(-at))$) для $k=1;5$, $a=0,5;2$; 2) з екстремумом (частина синусоїди $A\sin(\omega t)$ для $A=5$, $\omega=0,47$, $t=5c$).

4. Зміст роботи Вступ 1. Загальна характеристика завдань прогнозування процесів 2. Загальна характеристика нейронних мереж 3. Моделювання процесів та їх прогнозування з використанням нейронних мереж. 4. Порівняльний аналіз ефективності нейронних мереж для прогнозування процесів. Висновки

5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо) презентація _____

6. Консультанти розділів роботи*

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

7. Дата видачі завдання 20.04.2020

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1.	Огляд літератури з напрямку прогнозування процесів	30.04.20	
2.	Порівняльна характеристика нейронних мереж	10.05.20	
3.	Моделювання монотонних процесів та процесу з екстремумом	15.05.20	
4.	Обґрунтування структури нейронної мережі, розробка програмного забезпечення, виконання прогнозування змодельованих процесів	29.05.20	
6.	Дослідження ефективності прогнозування	05.06.20	
7.	Оформлення пояснювальної записки. Підготовка до захисту	09.06.20	

Студент

(підпис)

_____ А.О. Дорошенко

(ініціали, прізвище)

Керівник дипломної роботи

(підпис)

_____ Н.І. Бурау

(ініціали, прізвище)

* Консультантом не може бути зазначено керівника дипломної роботи.

ЗМІСТ

АНОТАЦІЯ.....	5
СКОРОЧЕННЯ, ПРИЙНЯТІ В ДИПЛОМНОЇ РОБОТІ.....	7
ВСТУП.....	8
1. ЗАГАЛЬНА ХАРАКТЕРИСТИКА ЗАВДАНЬ ПРОГНОЗУВАННЯ ПРОЦЕСІВ.....	9
1.1 Визначення та основні поняття.....	10
1.2 Прогнозування фінансових ринків.....	12
1.3 Прогнозування ринку електроенергії.....	14
1.4 Прогнозування поточного стану економіки.....	18
1.5 Прогнозування захворювань. Моделі поширення захворювань.....	20
1.6 Часові ряди.....	24
1.6.1 Структура математичної моделі.....	25
1.6.2 Авторегресійні моделі прогнозування.....	27
Висновок до розділу 1.....	29
2. ЗАГАЛЬНА ХАРАКТЕРИСТИКА НЕЙРОННИХ МЕРЕЖ.....	31
2.1 Персептрон.....	34
2.2 Мережі з радіальними базисними функціями.....	35
2.3 Мережі зі зворотним зв'язком.....	36
2.4 Рекурентні нейронні мережі.....	37
2.5 Довга короткострокова пам'ять (LSTM мережа).....	38
2.6 Методи і алгоритми навчання нейронних мереж.....	40
Висновок до розділу 2.....	43
3. ДОСЛІДЖЕННЯ ДЕТЕРМІНОВАНИХ ПРОЦЕСІВ.....	45
3.1 Лінійна нейрона мережа.....	45
3.2 Рекурентна нейронна мережа LSTM.....	54
Висновок до розділу 3.....	59
ВИСНОВКИ.....	60
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	61
ДОДАТОК А.....	63
ДОДАТОК Б.....	68

АНОТАЦІЯ

У даній роботі розглядається задача виділення методичного та алгоритмічного забезпечення в прогнозуванні фізичних процесів. Як інструмент прогнозу оглядаються різні типи нейронних мереж, які застосовуються до задачі прогнозування. Такий підхід доцільний у випадках, коли класичні статистичні методи не можуть в повній мірі задовольнити задачі прогнозу, а саме неможливість оцінити нестационарність, неповноту або невідомий розподіл даних.

ЧАСОВІ РЯДИ, МАТЕМАТИЧНА МОДЕЛЬ, НЕЙРОННІ МЕРЕЖІ

ABSTRACT

In this paper considered the problem of definition of methodological and algorithmic support in predicting physical processes. As a forecasting tool, different types of neural networks that are applied to the forecasting problem are considered. This approach is appropriate in cases where classical statistical methods cannot fully satisfy the task of forecasting, namely the inability to assess the non-stationarity, incompleteness or unknown distribution of data.

TIME SERIES, MATHEMATICAL MODEL, NEURAL NETWORKS

СКОРОЧЕННЯ, ПРИЙНЯТІ В ДИПЛОМНОЇ РОБОТІ

АКФ – автокореляційна функція

НМ – нейронна мережа

РБФ – радіально базисна функція

СКВ – середньоквадратичне відхилення

ШНМ – штучна нейронна мережа

ARMA (англ. autoregressive moving-average model) – авторегресивне ковзне середнє

ARIMA (англ. autoregressive integrated moving average) – інтегрована модель авторегресії ковзного середнього

LTSM (англ. Long short-term memory) – різноманітність архітектури рекурентних нейронних мереж

RNN – рекурентна нейронна мережа

ВСТУП

В епоху глобалізації і майже повсюдною загальнодоступності інформації, завдання прогнозування часових рядів залишається актуальною, особливо помітний великий стрибок в даному напрямку зважаючи на доступність все більш потужних засобів збору і обробки інформації. Прогнозування часових рядів – важлива науково-технічна проблема, що дозволяє передбачити поведінки різних систем і факторів в економічній, соціальній, екологічній, медичній сфері і в інших. На сьогоднішній день нейронні мережі знаходять застосування в різних галузях, починаючи від медичної та економічної, закінчуючи інформатикою та управлінням соціальними інститутами. Основний здатність нейромереживих технологій є здатність до обробки різнобічної інформації, здатності до виділення і класифікації залежностей між вхідними та вихідними даними.

РОЗДІЛ 1

ЗАГАЛЬНА ХАРАКТЕРИСТИКА ЗАВДАНЬ ПРОГНОЗУВАННЯ ПРОЦЕСІВ

1.1 Визначення та основні поняття

Основне завдання прогнозування - формування цілісної оцінки будь-якої системи в чітко визначений момент або проміжок часу і формування прогнозу поведінки цієї системи на підставі аналізу різних залежностей між різними факторами. Прогнозування будь-якого процесу будується на основі сукупності методів для розробки оптимальних алгоритмів процесу, що аналізується. Система прогнозу - організаційно-технічна система, що дозволяє отримати деякі дані про майбутній стан про об'єкту прогнозу, за допомогою даних, що надходять на вхід, тобто з метою отримання прогнозу. У якості такої системи надалі будуть розглядатись нейронні мережі [1].

В якості основної вимоги до досліджуваної системи є здатність до реагування на зміни, які можуть відбуватися в об'єкті дослідження. А отже можна виділити три основні критерії до досліджуваної системи [2]:

- Максимальне зменшення рівня невизначеності майбутньої ситуації.
- Швидке розпізнавання змін в закономірностях функціонування об'єкта і провадження подальших обчислень прогнозування з урахуванням цих змін.
- Здатність досліджуваної системи відрізняти зміни в об'єкті прогнозування від результату впливу невизначеності.

Основні вимоги до прогнозуючої системі можуть змінюватися в залежності від конкретного завдання. Наприклад, для систем прогнозування фінансових ринків основною необхідною вимогою є підвищена здатність фільтрації корисної інформації від "шуму".

В залежності від об'єкта дослідження, необхідне правильне визначення моделі поведінки даного об'єкта. Будь-яке прогнозування будується на основі інформації, отриманої в результаті спостереження. Відповідно, на основі статистичних даних існуючої інформації, можна виділити закономірності в поведінці об'єкта, і, як наслідок, є можливість дослідження, шляхом отримання його максимально наближеною математичної моделі. На вибір оптимальної моделі безпосередньо впливає мета і завдання прогнозування, а також інтервал, на який воно проводиться.

Основною вимогою до обраної моделі є необхідність врахування основних залежностей, що описують обраний об'єкт дослідження, а також фільтрації другорядних. Як наслідок, варто виділити основні проблеми при виборі математичної моделі: обсяг відповідної інформації про об'єкт дослідження, ступінь вивченості процесу, що спостерігається за досліджуваним об'єктом, а також наявність відомостей про некоректність наявній інформації [3].

Виходячи з отриманої інформації про об'єкт дослідження, математичні моделі умовно можна класифікувати наступним чином: аналітичні, статистичні та комбіновані [1].

Аналітичні (теоретичні) моделі, в основному, в процесі побудови ґрунтуються на основі відомих законів природознавства. Для побудови правильної аналітичної моделі необхідна достатня кількість спостережень і досліджень, тобто обраний об'єкт повинен бути добре вивчений. Також варто враховувати, що при побудові такої моделі використовуються деякі припущення і спрощення зважаючи на виключення другорядних параметрів, що безпосередньо впливає на точність моделювання і, як наслідок, прогнозування.

Статистичні (емпіричні) моделі базуються на підставі деяких статистичних даних, а також на підставі результату їх аналізу. Зважаючи на великий обсяг досліджуваної інформації, для таких моделей необхідний вибір правильного математичного виразу для більш точного прогнозування на підставі

експериментальних даних. Як наслідок, такого роду моделі найменш універсальні.

Комбіновані моделі, в свою чергу, здатні різною мірою поєднувати плюси вищенаведених моделей з огляду на зменшення систематичної похибки моделі, шляхом використання коефіцієнтів корекції.

Крім вищенаведених класифікацій, математичні моделі можна розділяти за такими ознаками [1]:

- За кількістю змінних - одномірні і багатовимірні моделі;
- За характером змін - динамічні і статичні;
- За характером зміни в часі - стаціонарні та нестаціонарні;
- За характером зв'язків - лінійні і нелінійні;
- За поданням в часі - безперервні і дискретні;
- За характером процесу - детермінований і стохастичний (випадковий).

Дослідження обраного об'єкта або процесу так чи інакше призводить до побудови прогнозуючої системи і, як наслідок, до вибору і аналізу математичної моделі. Тип математичної моделі залежить від природи спостережуваного об'єкта або процесу, а також від бажаного результату і завдання дослідження. Математична модель, в основі своїй, містить опис взаємозв'язків переходу об'єкта або процесу з одного стану в інше.

В залежності від мети і об'єкта дослідження, можна виділити основні методи побудови математичної моделі: Імітаційне, інформаційне та структурне моделювання [1].

Імітаційне моделювання має на увазі, що при побудові математичної моделі буде зроблена спроба максимально відтворити процес зміни об'єкта в часі за допомогою алгоритмів. В такому випадку імітуються основні взаємозв'язки, які цей процес утворюють. Такого роду моделювання дає можливість враховувати деякі ознаки об'єкта, наприклад, Дискретність або безперервність у часі, лінійність, характер зміни в часі.

Структурне моделювання базується на поділі і класифікації наявної інформації про об'єкт. Подальше моделювання таким методом ґрунтується на дослідженні окремих взаємозв'язків об'єкта, що спостерігається, шляхом використання інших методів.

1.2 Прогнозування фінансових ринків

Фінансові ринки - класичний приклад використання задач прогнозування. Дана сфера дуже популярна через великої кількості методів формування прогнозу часових рядів.

В аналізі функціонування фінансового ринку лежить концепція ефективного ринка. В основі даної концепції передбачається, що в структурі фінансового ринку сторони, що обмінюються, отримують вигоду від угод без оплати транзакцій [4].

Основна суть гіпотези - фінансовий ринок може вважатися ефективним в тому випадку, якщо інформація, що впливає на ринок, повністю і в найкоротші терміни відбивається на ціні активу. Дана концепція була сформована американським економістом Юджином Фамой в 60-х роках XX століття. Гіпотеза не може в повній мірі пояснити цінові рухи зважаючи на недоступність деякою інформації широкого кола учасників фінансового ринку. Частина такої інформації часто залишається приватною і, в першу чергу, доступна невеликому колу осіб. Тому очевидно, що така концепція не може існувати в чистому вигляді в реальному світі, і може застосовуватися тільки в якості основи зразкової моделі поведінки цінових активів.

Для більш широкого аналізу ринків, в основному, застосовується два типи його аналізу: технічний [5] і фундаментальний аналіз [7].

Технічний аналіз - набір інструментів аналізу руху цін активів, проводиться переважно на основі графіків і передбачає прогнозування

майбутньої динаміки цін на основі минулого руху. Іншими словами, аналіз на основі ринкової статистики.

Технічний аналіз базується на використанні [5] патернів, діаграм, інших моделей поведінки, для оцінки і прогнозування майбутніх цінових рухів. Такого роду аналіз може використовуватися як інструмент для прогнозу, наприклад, акцій, ф'ючерсів, ринку цінних паперів, форекс та інших високо волатильних активів. Основа технічного аналізу фінансового ринку була запропонована Чарльзом Доу [6] в кінці 1800-х років і була надалі розвинена шляхом багаторічних досліджень в даній області.

Для використання такого аналізу передбачається, що поточна ціна повністю відображає всю доступну інформацію про активі, тому його ціна справедлива і може бути базою наступного прогнозу. В основі лежить фіксована ціна в певні моменти часу, а, отже, наявну інформацію можна використовувати для розпізнавання тренду.

У технічному аналізі також передбачається, що ціни на активи можуть рухатися направлено і хаотично в різні моменти часу. А виходячи з припущення, що рух цін не випадковий і в певних моделях і тенденції можуть повторюватися, слідує, що ринкові тренди можуть бути ідентифіковані. А так як періоди руху цін можуть бути різними, то технічний аналіз можна проводити, як на короткострокові тренди, так і на довгострокові.

Як наслідок з вищесказаного, базової метою технічного аналізу лежить прогнозування тренду, лише на основі попередніх даних руху цін активів.

Фундаментальний аналіз - аналіз цін активів, який базується на оцінці якісних і кількісних факторів, тобто мається на увазі збір і аналіз інформації, що впливають на вартість активу. Такого роду аналіз заснований на зовнішніх подіях і впливах, тенденціях, фінансових звітів та іншому. А основне прогнозування базується на ключових показниках активу, бізнесу, галузі та визначенні його фінансового стану. На відміну від технічного аналізу,

фундаментальний аналіз розглядається як інструмент аналізу факторів поза руху ціни самого активу [7].

Всі існуючі методи і інструменти, які використовуються в фундаментальному аналізі можна розділити на дві частини: низхідному і висхідний. При низхідному методі мається на увазі, що аналіз буде проводитися, починаючи від всього ринку, включаючи глобальні макроекономічні показники, проходячи через конкретний сектор до конкретного активу. Іншими словами - оцінка загальної ситуації, з пошуком причинно-наслідкового зв'язку подій, що відбуваються, до приватного висновку. Висхідний метод, в свою чергу, є повною протилежністю низхідному [7].

Фундаментальний аналіз, на відміну від технічного, підходить для прогнозування тренду в більш довгостроковій перспективі. Для того аналізу менше значення має поточний рух цін на активи.

Для якісного прогнозування руху цін на фінансових ринках необхідно враховувати безліч супутніх чинників: наявність хаотичного руху цін в деякі періоди часу, відсутність повноти інформації для глобального аналізу, велика зашумленість вхідних даних, необхідність постійного удосконалення математичної моделі.

1.3 Прогнозування ринку електроенергії

Також досить популярною областю застосування нейронних мереж є прогнозування ринку виробництва і споживання електроенергії. Дана сфера діяльності має велике значення для світових і національних економік. Однак в сфері енергетики спостерігається ряд проблем: частина методів та інструментів прогнозу не враховують якісні зміни на ринку.

Електроенергія в якості товару споживання, як і ціна активу на фінансовому ринку, відображає сукупність характеристик, що включають в себе

основний фактор попиту і пропозиції, тобто виробництва і споживання. І аналогічно фінансових ринків, на ринок електроенергії також впливають безліч зовнішніх чинників, наприклад, споживання в різні години дня, день тижня, сезонність споживання, погодні умови та інші.

Ринок електроенергії є монопродуктовим, товар якого - електроенергія. Для прогнозу ринку енергоресурсів можна виділити три основні агрегати: ціна енергоресурсу, обсяги виробництва і попиту. На відміну від двох останніх показників, які схильні до монотонним тенденціям і підлягають прогнозом, ціни, ціна - високо волатильний показник і, отже, важко піддається довгостроковим прогнозам [8].

Для аналізу ринку електроенергії необхідна побудова моделі його функціонування з урахуванням фізичних властивостей електроенергії. Слід виділити наступні властивості [8]:

- На довгий період часу неможливо створити достатньої кількості запасів електроенергії;
- Прогнозування і відстеження залежностей споживання деяких суб'єктів системи;
- Прогнозування обсягів виробництва на довгий період часу неможливо з великою точністю;
- Кількість виробництва в належній мірі має збігатися з кількістю споживання.

Отже, основною метою ринку електроенергії є баланс між виробництвом і споживанням. І, як наслідок, можна виділити основну мету прогнозу: довгостроковий прогноз стану суб'єктів споживання, що в свою чергу впливає на цінову політику ринку.

Варто відзначити, що в реальних умовах неможливо досягнення повного балансу зважаючи на присутність відхилень між споживанням і виробництвом, заснованих на зовнішніх факторах, наприклад: зміни в погодних умовах,

економічні, політичні зміни. Дані зміни не можуть бути спрогнозовані на довгий період часу, отже і довгостроковий прогноз ринку електроенергії, з урахуванням цих змін, не може бути проведений з належною точністю.

Безпосередній вплив сезонних коливань, а також короткострокових змін в споживанні, примушує виробника до збільшення виробництва електроенергії, що призводить до невживаних залишків. У свою чергу, неможливість створення запасів змушує до утворення резервів виробничих потужностей. Витрати на збільшення виробництва або на утримання запасів безпосередньо впливає на підвищення кінцевих цін електроенергії.

Основні методи аналізу та прогнозування ринку електроенергії можна розділити на п'ять типів [9].

Мультиагентний метод - метод імітаційного моделювання, заснований на теорії ігор. Дана модель базується на імітації роботи агентів ринку, будь то окремих осіб або колективних суб'єктів, таких як організації або групи. Кінцевою метою такого моделювання є оцінка впливу агентів на системи і ринок. Мультиагентний метод являє собою набір гнучких інструментів для аналізу поведінки на ринках електроенергії. Однак дана свобода є і мінусом зважаючи на підвищені вимоги до обґрунтування припущень, закладених в моделі.

Фундаментальний метод - метод, в основі якого лежить аналіз впливу фізичних та економічних факторів. Для такого методу істотні дві проблеми: доступність даних і включення стохастичних коливань. Отже, через збір інформації, який, як правило, відбувається за тривалі терміни, і наявності конкретних припущень про фізичних та економічних припущеннях на ринку, фундаментальний метод більше підходить для середньострокових прогнозів, ніж для короткострокових, а також досить чутливий до раніше висунутих припущень про закономірності ринку.

Моделі зменшеної форми - метод, який базується на статистичних показниках і особливості торгівлею електроенергією, - оцінка та управління

ризиками. Для таких моделей найважливішим є вибір цінового процесу. Якщо обраний цінової процес не підходить для визначення основних властивостей цін на електроенергію, результати моделі, швидше за все, будуть ненадійними. Такі моделі надають спрощену, але досить реалістичну інформацію про динаміку цін.

Статистичні моделі - моделі, які містять часові ряди і економетричні показники. Даний метод за своєю структурою схожий на метод технічного аналізу фінансових ринків. Статистичні методи прогнозують поточну ціну, ґрунтуючись на комбінації попередніх цін і зовнішніх чинників. А точність безпосередньо залежить від якості використовуваних алгоритмів і здатності обліку зовнішніх факторів, наприклад історичні прогнози попиту і споживання, погодні умови та інші. Однак даний метод працює досить погано при наявності піків.

Методи штучного інтелекту. Метод об'єднує елементи навчання, еволюції та нечіткості, щоб створити підходи, які здатні адаптуватися до складних динамічних і багато параметричних систем. Такі моделі є гнучкими і можуть справлятися зі складністю і нелінійністю. Головним достоїнством і одночасно і недоліком методу штучного інтелекту є їх здатність справлятися з складністю і нелінійністю. В цілому, такий метод краще моделює особливості поведінки цін, ніж статистичні. У той же час, їх гнучкість не обов'язково призводить до кращих точковим прогнозами.

Ринок електроенергії є частковим випадком фінансового ринку. Для обох областей дослідження виникає проблема врахування зовнішніх факторів. Однак у випадку електроенергії, необхідно враховувати не тільки ціну і зовнішні макроекономічні фактори, а й фізичні властивості електрики як товару. Дана проблема частково вирішується наявністю основних методів аналізу. Але, варто визначити, що використання конкретного методу побудови математичної моделі має враховувати в максимальній мірі характеристики розглядаємого процесу, що залежать від мети прогнозування. А отже, доцільно використовувати різні способи одночасно.

1.4 Прогнозування поточного стану економіки

Як більш глобальний випадок прогнозування в сфері економіки, варто також виділити окремо прогноз економічного курсу. Для деяких з вищенаведених прикладів істотна проблема впливу фактора поточного і майбутнього стану економічних показників. Аналіз, який базується на дослідженні кількісних показників об'єкта спостереження, так чи інакше, намагається виділити закономірності впливу зовнішніх факторів. Однак, дані, які беруться до розгляду, в основному, агреговані або представляються такими. Тому для більш якісного аналізу і прогнозування систем, які отримують безпосередній вплив від стану економіки, варто виділити окремо прогнозування даних економічних показників. Створення і використання ефективних систем аналізу економіки і її прогнозування сприятливо позначиться на вирішеннях завдань прогнозу, не тільки, які залежать від економічних показників об'єктів дослідження, а й для прикладного вирішення самих економічних проблем, що, в свою чергу, приведе до підвищення її ефективності [10].

Існує кілька підходів до аналізу та прогнозування економічних показників.

Системний підхід - підхід до вивчення, в якому економічний процес розглядається як система, що складається з окремих елементів, які взаємопов'язані між собою і можуть впливати, як один на одного, так і на систему в цілому. Кожен елемент такої системи надає відмінні характеристики даного процесу.

Історичний підхід - підхід оцінки динаміки розвитку процесу. Дослідження взаємозв'язків, які привели цей процес в поточний стан, дозволяє виявити закономірності, тенденції розвитку в майбутньому [10].

Альтернативний підхід - передбачає використання декількох методів при побудові прогнозу.

Диференціальний підхід - передбачає використання специфіки об'єкта дослідження.

Математичний підхід - передбачає використання математичних методів прогнозування і способів їх побудови. Зокрема, використання статистики та нейронних мереж.

Аналіз економічних показників розділяється на два великі розділи за ступенем формалізації: інтуїтивний і формалізований; які діляться за загальним принципом дії і способу отримання прогнозованої інформації.

Інтуїтивні методи використовуються в разі, коли неможливо провести кількісну оцінку окремих процесів або в разі неможливості обліку безлічі факторів через складність досліджуваного процесу або об'єкта. Інтуїтивні методи базуються на використанні експертних оцінок, за допомогою яких встановлюються важливі фактори і взаємозв'язки між ними, визначаються критерії. У свою чергу, такі методи поділяються на два основні класи, а саме на індивідуальні та експертні.

Формалізовані методи діляться на два види: метод екстраполяції тенденцій, метод моделювання. Методи моделювання, в свою чергу, діляться на структурний, імітаційне, мережеве та статистичне моделювання, окремі випадки методів моделювання були описані раніше в основних методах аналізу і прогнозування ринку електроенергії. Методи екстраполяції тенденцій базуються на допущенні того, що фактори, що впливають на об'єкт дослідження, незмінні і визначають стан об'єкта в майбутньому на основі закономірностей в минулому.

Аналіз поточного і майбутнього стану економіки займає важливу частину, як і в прогнозуванні цін активів фінансових і електроенергетичних ринках, так і в прогнозуванні макроекономічних показників в цілому. Також визначення економічного курсу відіграє велику роль у складанні правильної математичної моделі обраного часового ряду. Економіка - сукупність відносин в системі виробництва, розподілу, обміну та споживання. Тому глобальне прогнозування

економічних показників має включати в себе врахування максимальної кількості факторів, а також взаємозв'язків між ними. Для завдання прогнозування, як варіант вирішення, доцільно використовувати нейронні мережі через їх можливості врахування та класифікації взаємозв'язків незримих факторів.

1.5 Прогнозування захворювань. Моделі поширення захворювань

Епідеміологічні прогнози є соціально-економічно важливими для держави. А для вивчення закономірностей поширення хвороби, виявлення причин та умов виділяється окрема роль в охороні здоров'я.

В залежності від необхідної мети, епідеміологічні прогнози виконуються на різні терміни. У той час, як короткострокове прогнозування виконується для оперативного управління і виявлення спалахів захворюваності, середньострокове прогнозування, в свою чергу, використовується для тактичного управління і прийняття рішення для підготовки до можливого спалаху. Побудова точного довгострокового прогнозу в належній мірі зробити неможливо, однак таке прогнозування має стратегічну важливість у підготовці персоналу, контроль виробництва медичних препаратів і так далі.

Виділяють кілька основних категорій епідеміологічного прогнозування: статистичні методи, методи математичного моделювання поширення інфекції, методи на основі машинного навчання.

Основними моделями статистичних методів є регресійні і авторегресійні моделі. Регресійна модель [11] є класичним і одним з найпопулярніших методів прогнозування. Основне завдання регресійного аналізу базується в знаходженні залежності між невідомими параметрами і в формуванні залежності між захворюваністю і факторами, які на неї впливають. Також такі моделі прийнято розділяти на адаптивні і не адаптивні. Для не адаптивних моделей використовуються всі наявні дані про об'єкт дослідження, в свою чергу,

адаптивні моделі розраховуються на основі ковзного вікна - формування навчального набору даних з членів часового ряду, який використовується при побудові моделі.

Авторегресійна модель [11] - модель, в якій поточне значення залежить лінійно від попередніх значень цього ж часового ряду, тобто поточне значення дорівнює зважений сумі кількох попередніх значень ряду. Одним з часто використовуваних моделей авторегресії є модель Бокса-Дженкінса - розширена модель ковзного середнього. Основний підхід в аналізі таким методом передбачає в першу чергу оцінювати стаціонарність ряду.

Обидва вище представлених методу аналізу - базові, підходять до аналізу будь-яких часових рядів і прогнозування на будь-який термін. Однак такі моделі в більшості випадків ігнорують локальні коливання показників, тому істотна проблема прогнозування на середній або довгий термін з належною точністю.

Альтернативою статистичним методом можуть бути методи машинного навчання. Штучні нейронні мережі широко застосовуються в рішенні задач, аналітичне рішення яких важко. Основним завданням у використанні штучних нейроммереж, як в медицині, так і в інших наукових сферах, є навчання штучна нейронна мережа (ШНМ), яке полягає в обчисленні вагових коефіцієнтів зв'язків. Мета навчання - виявлення явних і неявних зв'язків між вхідними та вихідними даними. З огляду на зашумленність вхідних даних, для навчання штучних нейронних мереж потрібна велика кількість вихідної інформації для аналізу, наприклад, статистики захворюваності.

Байєсовські мережі [12] - тип нейронних мереж, який в своїй структурі враховує кореляції даних. Вагові коефіцієнти такої мережі розраховуються на основі імовірнісних залежностей, які задані функціями розподілу. Побудова таких мереж може відбуватися при різному обсязі наявної інформації, однак, зважаючи на складність обчислення параметрів оцінки, байєсовські мережі надають можливість тільки для короткострокового прогнозу захворюваності.

Перераховані вище техніки відносяться до загальних методів аналізу часових рядів і не враховують специфічні властивості епідеміологічних процесів. Аналіз ґрунтується на видимих показниках, статистики рівня захворюваності, і може застосовуватися при прогнозуванні будь-яких інших процесів. Методи, що базуються на явному моделюванні механізму передачі інфекції, поділяють на три основні категорії: аналітичне та імітаційне моделювання, а також їх комбінований тип.

В основі аналітичного моделювання лежать системи диференціальних рівнянь, які описують динаміку поширення інфекції. В якості основної моделі використовують модель SIR. У таких моделях досліджувану групу людей ділять на три головні підгрупи: здорові (S - susceptible), хворі (I - infectious) і видужали людей (R - recovered). Диференціальне рівняння набуває наступного виду [12]:

$$\frac{dS(t)}{dt} = -\alpha S(t)I(t), \quad \frac{dI(t)}{dt} = \alpha S(t)I(t), \quad \frac{dR(t)}{dt} = \beta I(t) \quad (1.1)$$

Де α відображає ймовірність передачі інфекції, а $1/\beta$ - тривалість заразливого періоду. Мета використання SIR - моделювання контактів в групі людей, наближене максимально можливо до реального. Також можна виділяти людей з захворюванням в інкубаційному періоді (E - exposed), вакцинованих індивідів (V - vaccinated), вікові групи населення, якщо цього вимагає завдання моделювання і особливості поширення інфекції. SIR моделі припускають наявності модифікацій, наприклад, із застосуванням динамічних параметрів α і β . А додавання стохастичних параметрів виправляє неможливість визначення імовірнісних процесів захворюваності. Також важливим аспектом моделювання епідемій є фактори розподілу інфекції: вплив географії, щільність населення, стан медичного апарату, рівень медичного страхування та інші. В якості вирішення в моделюванні методом SIR прийнято розділяти досліджувану групу людей на додаткові підгрупи. Перевага таких моделей у можливості їх аналітичного дослідження. Однак допускається, що всі характеристики підгрупи людей між собою однакові.

Найбільш популярним методом інтуїтивного моделювання є мультиагентний метод. Даний метод аналогічний вищеописаному, який застосовується в моделюванні поведінки учасників фінансового ринку. Моделювання поширення захворювань зводиться до опису оточення досліджуваної групи людей. Метод не використовує єдиного розрахунку ймовірності передачі інфекції, а описує поведінку кожної людини. Основоположним є моделювання подій, пов'язаних з групою людей, а отже, процес зводиться до послідовної обробці подій, де кожен агент формує таку обставину, яка пов'язана з його діяльністю, і, отже, впливає на систему в цілому. Такий метод максимально наближений до реальності, однак вимагає великої кількості обчислювальних потужностей.

Комбінований метод передбачає використання різних технік, модифікацій і комбінацій. Поширеними техніками прогнозування рівня захворюваності є техніки декомпозиції і усереднення часових рядів, фільтрація вхідної інформації або часового ряду. Різні поєднання залежать безпосередньо від розв'язуваної задачі і доречні, якщо використання інших підходів не дає належного результату.

Моделювання поширення захворювань дозволяють сформулювати прогноз на будь-який термін, однак побудова і аналіз моделі значно складніше інших сфер прогностики. Розробка математичної моделі і його настроювання ведеться для конкретного завдання, особливості якої включені в модель, а отже, одна модель поширення захворювань не може бути застосовна до іншої з отриманням результату належної точності. Як підсумок, варто підкреслити, що немає універсального способу прогнозування захворюваності. А оптимальний підхід необхідно вибирати, порівнюючи результати, отримані за допомогою різних методів.

1.6 Часові ряди

Моделі, які побудовані на основі статистичних даних, найбільш часто використовуються в прогнозуванні динаміки різних процесів. Крім фінансових, соціально економічних, медичних наукових областях, такі моделі можуть бути у прогнозуванні та обробки фізичних процесів, застосовні в менеджменті ризиків, формування бюджету країни, системах управління підприємств і інших.

На даний момент існує величезна кількість методів прогнозування лінійних і нелінійних процесів на основі використання часових рядів, основними з них виділяють наступні: авторегресійні моделі і їх види, штучні нейронні мережі, байєсовські мережі та інші види нейронних мереж, а також види моделей властиві певній галузі дослідження. Кожна запропонована модель може мати, як і плюси, так і мінуси, наприклад, основною перевагою нейромереж є їх точність, але, в той же час, високоточний прогноз вимагає великої обчислювальної потужності.

На основі наявних статистичних даних виділяють два основних типи моделювання та прогнозування: часові ряди і часовий розріз [13].

Часові ряди являють собою безліч, найчастіше, рівновіддалених між собою часу спостережень, які визначають і характеризують поведінку спостережуваного процесу. Часовий розріз - безліч спостережень за вибраний проміжок часу.

Часові ряди - основа аналізу поведінки різних систем. Наприклад, щоденні спостереження зміни фінансових активів, рівня електроспоживання та інші. Часові ряди можуть бути обмежені за своїм розміром для економічних систем, в той же час, часові ряди технологічних процесів можуть бути практично нескінченними.

Період дискретизації - величина, що безпосередньо впливає на характеристику часового ряду, повинна бути строго постійною. Визначення

періоду дискретизації можна визначити, наприклад, за теоремою Котельникова-Шеннона [13].

У свою чергу, для визначення f_{max} - максимальної частоти гармонійної складової, можна скористатися розкладанням в ряд Фур'є, або іншого способу ортогонального перетворення даних. Також, в окремому випадку, частоту дискретизації можна визначити виходячи з фізичних властивостей і особливостей досліджуваного процесу. Варто відзначити, що при збільшенні періоду дискретизації можна втратити необхідні характеристики процесу, а при невеликих значеннях може з'явитися лінійна залежність між вимірами. Тому вибір значення періоду дискретизації повинен бути максимально компромісним.

При побудові математичних моделей, метою яких є прогнозування, в першу чергу необхідно визначити термін прогнозу: короткостроковий, середньостроковий або довгостроковий. Як правило, прогнозування значень змінних виконується на базі більш простих моделей, ніж сам досліджуваний процес. Наприклад, дискретні моделі на основі авторегресійної моделі або її модифікації. Прогнозування може бути ускладнено наявністю нелінійних або нестационарних випадкових процесів. В такому випадку доцільно використання альтернативних методів або нейронних мереж.

1.6.1 Структура математичної моделі

Для побудови математичної моделі необхідно виділити її важливі параметри: вимірність моделі, порядок, наявність нелінійностей, характер нелінійності, якщо така присутня, час запізнення, тип збуджень.

Вимірність моделі - кількість рівнянь, що використовуються для математичного опису процесу [13].

Порядок моделі - порядок рівняння, що описує динаміку зміни процесу.

Нелінійність і її характер визначається в залежності від обраної моделі при її побудові, зустрічаються нелінійності щодо змінних або щодо параметрів.

Час запізнення - час реакції на виході щодо вхідних даних. Час запізнення чисельно дорівнює числу періодів дискретизації, на яке відстає вихідний сигнал від вхідного.

Тип збуджень - вхідні дані, які впливають на процес, і, як правило, створюють негативні умови його зміни. Збудження можна розділити на детерміновані і стохастичні, які, найчастіше, описуються розподілом випадкових величин.

Вибір структури моделі має першочергову важливість при побудові математичної моделі, а правильність вибору структурної моделі оцінюється можливістю моделі повторити основні змінні реального процесу.

Розділяють два основні підходи побудови математичних моделей: функціональний і структурний [13]. Структурний підхід має на увазі моделювання внутрішніх взаємодій змінних процесу і відображає їх фактичні взаємозв'язку. Під функціональним підходом мають на увазі формальний опис поведінки процесу без моделювання взаємодій його змінних. Плюсом такого методу в його відносній простоті і швидкості побудови математичної моделі.

Важливим фактором у правильності побудови моделі є вибір статистичних даних і оцінка параметрів моделі. Виділяють наступні вимоги обраних даних:

1. Дані повинні вимірюватися через однакові проміжки часу. В іншому випадку буде відбуватися порушення взаємозв'язків. Дані, які вимірюються не через рівні проміжки часу не підходять для побудови передатних функцій.
2. Для побудови моделі, вибірка даних повинна включати період часу досить довгий, який передбачає всі стану досліджуваного процесу або об'єкта.

3. Вибірка даних повинна бути інформативною. Оцінка відбувається за допомогою підрахунку кількості похідних, які можна взяти з вимірів. Або оцінку інформативності можна визначати величиною дисперсії. Вважається, що чим більше похідних можна взяти з вибірки або чим більше дисперсія, тим більше інформативна вибірка.

Безпосередня точність оцінки параметрів моделі залежить від якості вибірки даних, а також точності роботи і правильності вибору математичної моделі. Класичними вимогами до оцінки параметрів моделі наступні:

1. Оцінки не повинні мати систематичної помилки, яка збільшує або зменшує оцінку параметрів на вибірках даних або на певних відрізках.
2. Оцінка векторів параметрів повинна бути наближена до свого істинного значення в міру збільшення обсягу вибірки - повинна бути консистентним.
3. Оцінки повинні бути ефективними - значення середнього відхилення повинні бути найбільш наближеними до значення незміщене і консистентних оцінок.

1.6.2 Авторегресійні моделі прогнозування

Авторегресійні моделі і їх типи широко використовуються для опису стаціонарних випадкових процесів. Тобто випадковий розподіл функції залежить тільки від попередніх значень і відстані часу між значеннями. Модель, значення ряду якої визначаються як лінійна функція, називається авторегресійною (AR) [14].

Авторегресійний процес порядку p має наступний вигляд:

$$X(t) = c + \sum_{i=1}^p \varphi_i X_{t-i} + e(t), \quad (2.1)$$

де φ_i - параметри моделі, c – константа, $e(t)$ – білий шум.

Особливістю таких процесів є плавно загасаюча автокореляційна функція (АКФ).

Авторегресійна модель ковзного середнього (ARMA - autoregressive moving average) - більш компактна модель, що описує властивості стаціонарного часового ряду [14]:

$$X(t) = c + \sum_{i=1}^p \varphi_i X_{t-i} + e(t) - \sum_{i=1}^q \theta_i e(t-i), \quad (2.2)$$

де φ_i - коефіцієнти авторегресії, $e(t)$ - шумова складова, θ_i - коефіцієнти ковзного середнього, p, q - параметри моделі.

Інтегрована модель авторегресії змінного середнього (ARIMA - autoregressive integrated moving average) - Модель Бокса-Дженкінса [14]. Така модель є узагальненою простіший авторегресійної ковзної середньої і поняттям інтеграції, і є розширеною моделлю ARMA для нестационарних рядів. Такі часові ряди можна зробити стаціонарними завдяки взяття різниць деякого порядку від вихідного сигналу часового ряду. Така модель визначається наступним чином:

$$\Delta^d X_t = c + \sum_{i=1}^p \alpha_i \Delta^d X_{t-i} + \sum_{j=0}^q b_j e_{t-j}, \quad (2.3)$$

де α_i - коефіцієнти авторегресії, b_j - коефіцієнти ковзаючого середнього, c - константа, e_t - білий шум. Δ^d - оператор різниці часового ряду порядку d (послідовне взяття d раз різниць першого порядку - спочатку від часового ряду, потім від отриманих різниць першого порядку, потім від другого і т.д.)

Модель ARMA ($p + d, q$) визначається як модель ARIMA, при цьому, якщо $d = 0$ - інтерпретується звичайна ARMA модель.

Першочергову важливість використання методу ARIMA - оцінка стаціонарності ряду. В першу чергу визначається наявність поодиноких коренів і порядок інтегрованості часового ряду. Якщо порядок інтегрованості ряду більше нуля, то ряд необхідно перетворити шляхом взяття різниці відповідного порядку і потім для перетвореної моделі будується відповідна ARMA модель. В даному випадку мається на увазі, що побудований процес стаціонарний.

ВИСНОВОК ДО РОЗДІЛУ 1

Резюмуючи вищесказане, сфера економіки є класичним прикладом застосування прогнозування різними методами. В якості основної проблеми можна виділити наступне: велика кількість зовнішніх факторів, що впливають на процес зміни об'єкта дослідження, а також велика кількість параметрів, які необхідно враховувати в аналізі і прогнозу, не дають можливості провести точне прогнозування об'єкта економіки. З розвитком технологій і повсюдним обміном доступною інформацією між фахівцями різних наукових сфер діяльності, відбувається постійне збільшення кількості різних підходів, методів, інструментів аналізу і прогнозування. Доступним і популярним рішенням є нейромережеві технології. Штучні нейронні мережі являють великий інтерес фахівців з урахуванням їх ефективності та можливості нелінійного моделювання в поєднанні з порівняно простою реалізацією. Другою важливою властивістю, якою володіють нейромережі - здатність до навчання та узагальнення добутих знань. І як наслідок, натреновані нейронні мережі здатні швидко аналізувати велику кількість даних і узагальнювати отриману інформацію, при цьому показуючи гарні результати на даних, які не використовувалися при навчанні.

Незалежно від об'єкта дослідження, для якісного, інформаційно ємного, найбільш точного прогнозу слід виділити важливі переваги нейронних мереж в аналізі часових рядів. Стандартні статистичні методи не завжди здатні до визначення прихованих залежностей, що впливають на досліджуваний об'єкт або процес. Наявність великої кількості "шуму" у вхідних даних призводить до необхідної початкової фільтрації наявної інформації. Також, статистичні методи не можуть в повній мірі враховувати всі зміни в досліджуваному процесі через наявність нестационарності часового ряду або наявності випадкових величин. У свою чергу, нейромережеві технології здатні частково нівелювати вищеописані фактори. Отже, використання нейромереж доцільно в тому випадку, коли

класичні статистичні методи не можуть впоратись з поставленим завданням або не задовольняють очікуваному результату.

Підсумовуючи розділ, можна підкреслити, що для прогнозування з використанням часових рядів необхідна коректна побудова математичної моделі, досліджуваного процесу. Правильність моделі визначається вищеописаними вимогами і має на увазі їх обов'язкового виконання. У разі невиконання основних критеріїв, сформований прогноз не може бути максимально точним. Безпосередній вибір математичної моделі базується на, в першу чергу, цілі аналізу та прогнозуванні часового ряду, а по-друге, специфічних властивостях і характеристиках процесу дослідження.

Часові ряди можуть застосовуватися в аналізі за допомогою класичних методів прогнозування, а також за допомогою нейронних мереж, зважаючи на особливості розглянутих рядів: синхронність і інформативність наявних даних.

У разі, якщо зміна процесу імовірно лінійна, використання класичних методів більш пріоритетна зважаючи на їх відносну простоту і малої вимогливості до обчислювальної потужності техніки. Для нелінійних процесів - нейронні мережі можуть надати більш точний прогноз, однак вимоги до обчислювальних потужностей більш серйозні, а швидкість правильної побудови нейронних мереж і їх навчання значно нижче.

РОЗДІЛ 2

ЗАГАЛЬНА ХАРАКТЕРИСТИКА НЕЙРОННИХ МЕРЕЖ

Нейронна мережа- сукупність нейронів (елементів мережі), які з'єднані між собою синаптичними зв'язками - ваговими коефіцієнтами. Нейрони взаємодіють між собою завдяки наявності цих зв'язків. Спосіб їх з'єднання визначає архітектуру мережі. Найпростішим прикладом нейронної мережі є штучний нейрон, який імітує роботу біологічного нейрона [15].

Вхідний сигнал або деякий безліч сигналів подається на вхід нейрона, кожному вхідному сигналу

Вхідні величини множаться на відповідний ваговий коефіцієнт, добутки підсумовуються, а сума добутків визначає рівень активації нейрона. Сукупність вхідних сигналів позначається вектором X , безліч вагових коефіцієнтів позначається вектором W . Підсумовуючий блок складає вхідні значення, створюючи вихід NET , у векторному вигляді [15]:

$$NET = X * W \quad (2.1)$$

Вихідний сигнал NET , як правило, перетворюється функцією активації F і дає вихідний сигнал нейронної мережі OUT . Функція активації може бути лінійної або нелінійної. Якщо ж функція активації стискає вхідний сигнал NET , то таку функцію називають «стискаючою». В якості такої функції, найчастіше, використовують сигмоїдальну функцію:

$$f(x) = \frac{1}{1 + e^{-NET}} \quad (2.2)$$

Другою, також часто використовуваною функцією, є гіперболічний тангенс $th(x)$. Хоч дана функція і схожа зовні з сигмоїдальною, головна відмінна риса - симетричність щодо початку координат. А значить в точці $NET = 0$, вихідний сигнал OUT також буде дорівнювати нулю. Також важливу роль відіграє можливість даної функції активації приймати негативні значення, що

може бути корисно для деяких нейронних мереж. Обидві функції зображені на рис. 2.1 [15].

Штучний нейрон - найбільш спрощена модель поведінки біологічного нейрона. Для таких моделей характерна відсутність затримки в часі - тобто вхідні сигнали моментально породжують вихідний. Не дивлячись на простоту моделі штучного нейрона, така мережа вже може справлятися з найпростішими завданнями прогнозування або класифікації.

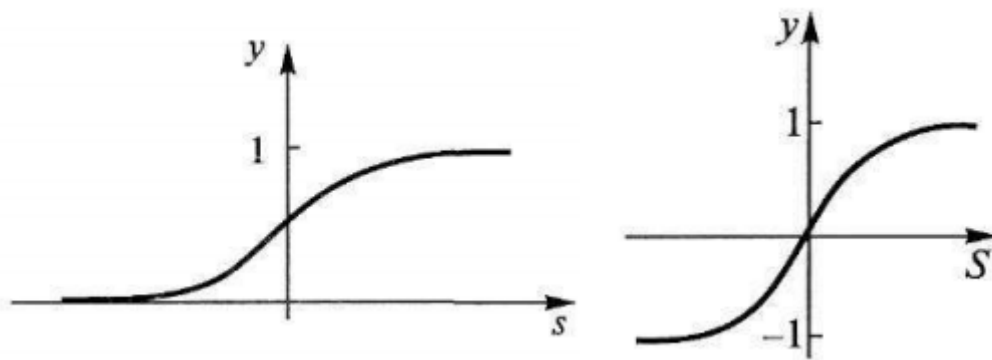


Рис. 2.1 Функції активації нейронів: а – сигмоїдальна, б – гіперболічний тангенс

Найпростіша нейронна мережа ґрунтується на обчисленнях, яке відбувається в з'єднаннях нейронів, і складається з групи нейронів, що утворюють шар. Схема одношарової нейронної мережі зображена на рис. 2.2 [15].

Кожен елемент вхідного сигналу X з'єднаний з кожним штучним нейроном за допомогою вагового коефіцієнта. Виходом такої мережі буде послідовність значень зваженої суми нейронів. Вихідний вектор розраховується аналогічно штучному нейрону, за винятком того, що матриця ваг W складається з m рядків і n стовпців.

Таким чином вихідний вектор для N і X векторів-рядків можна записати як [15]:

$$N = XW \quad (2.3)$$

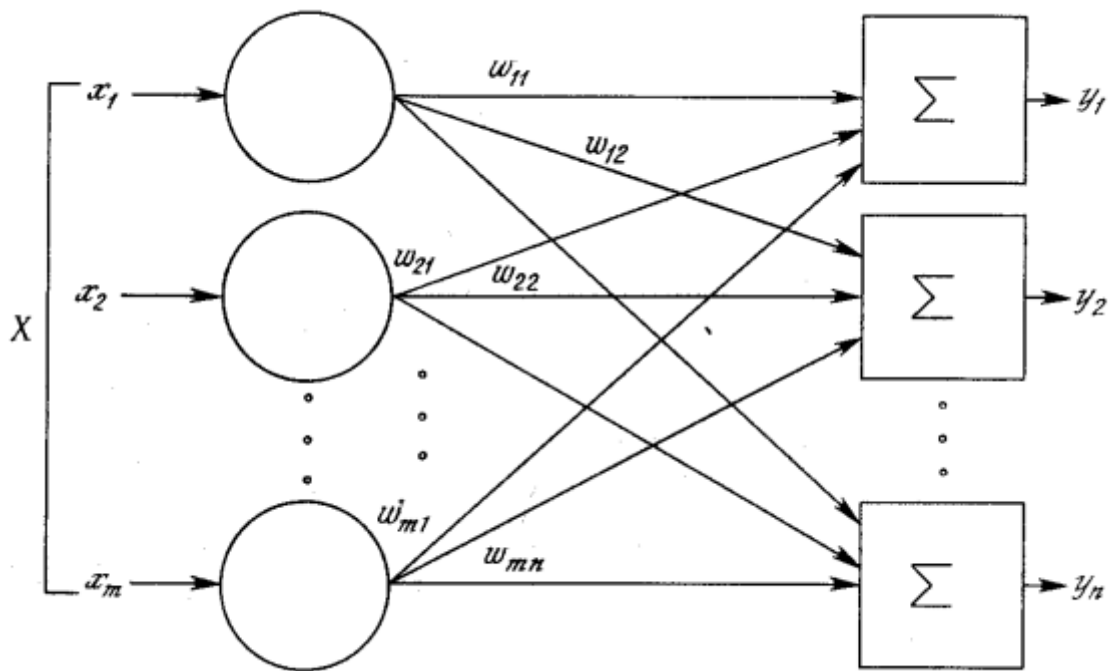


Рис. 2.2 Одношарова нейрона мережа

Багатошарові нейронні мережі (НМ) утворюються каскадами шарів, де вихід одного шару є входом для другого. Двошаровий нейронна мережа може бути представлена як одношарова, у якої з матриця вагових коефіцієнтів дорівнює добутку першої матриці W_{mn} і другий W_{mp} . Структурна схема багатошарової мережі зображена на рис. 2.3 [15].

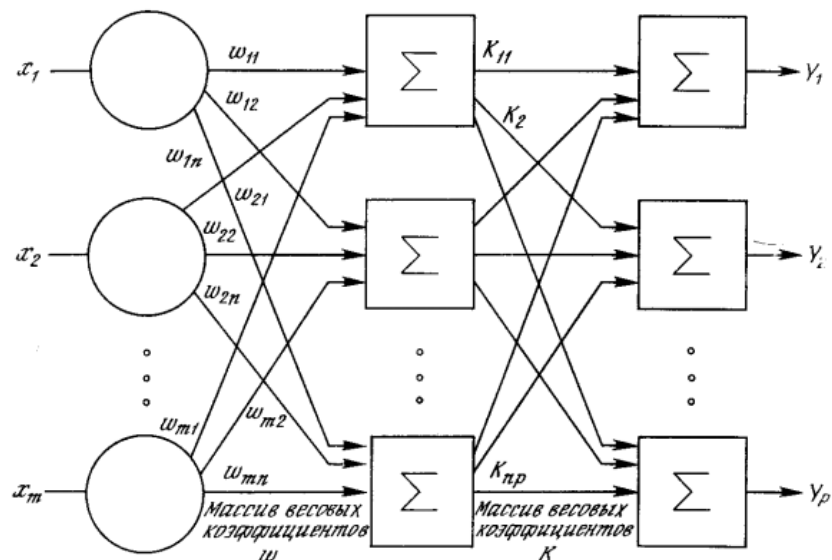


Рис. 2.3 Багатошарова НМ

Багатошарові нейронні мережі можна розбити на два підкласу [15]: НМ із зворотними зв'язками і НМ без зворотних зв'язків. Мережі без зворотних зв'язків називають мережами прямого поширення. Мережі, які мають з'єднання від виходів до входів, називають мережами з зворотними зв'язками. Для таких мереж характерна передача вихідного значення на вхід, а значить наступний вихід буде визначатися одночасно поточними значеннями на вході і попередньому виході.

2.1 Персептрон

Модель персептрона схожа з моделлю штучного нейрона. Персептрон складається з одного шару штучних нейронів, з'єднані ваговими коефіцієнтами. Кожен елемент входу множиться на ваговий коефіцієнт, отримані значення підсумовуються. Відмінною рисою персептрона є його функція активації, яка представляє жорсткий поріг. Якщо отримана сума більше заданого порогу, то вихід дорівнює одиниці, інакше нулю. Схема персептрону та його функція активації зображена на рис. 2.4 [15] та рис. 2.5 [15] відповідно.

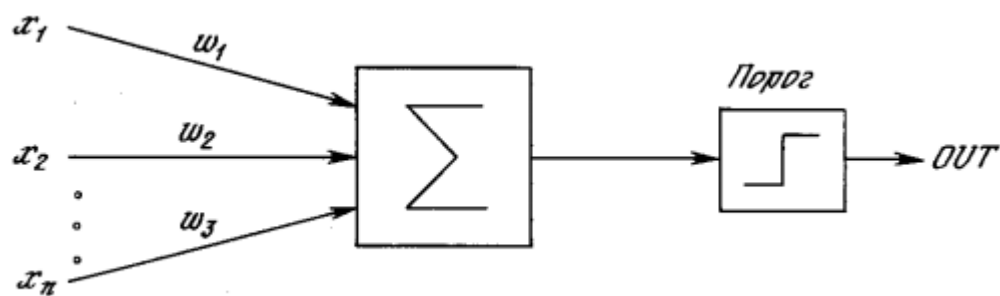


Рис. 2.4 Схема персептрону

Функція Хевісайда у якості функції активації дозволяє застосовувати персептрон в задачах класифікації, проте до завдань прогнозування не підходить. Персептрон складається з трьох типів елементів: сигнали, що надходять від

сенсорів, передаються до асоціативних елементів, а потім реагуючим елементам. Так, персептрони дозволяють створити набір «асоціацій» або зв'язків між вхідними сигналами і очікуваною реакцією на виході. Відповідно до сучасної класифікації [15], персептрони можуть бути віднесені до нейронних мереж з одним прихованим шаром, з порогової функцією передачі, з прямим розповсюдженням сигналу. Тому, в задачах прогнозу і визначення тренда слід замінити функцію активації на лінійну або будь-яку іншу в залежності від поставленого завдання і розглянутого процесу.

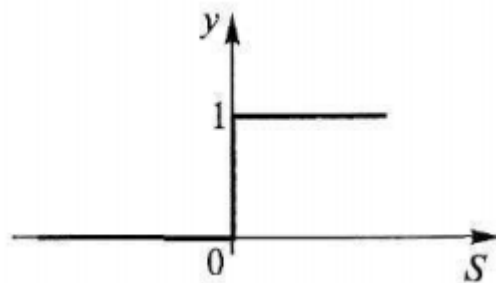


Рис. 2.5 Порогова функція активації персептрону (функція Хевісайда)

2.2 Мережі з радіальними базисними функціями

Нейронна мережа радіальних базисних функцій, в загальному випадку, містить в собі три шари: звичайний вхідний шар, шар прихованих нейронів з радіально базисної функцією, вихідний шар [15].

Такі НМ мають ряд переваг над іншими багатошаровими нейронними мережами. Можливість моделювання довільної нелінійної функції за допомогою одного шару оптимізує побудову мережі, а кількість можливих шарів зменшується. Також такі мережі можуть бути оптимізовані за допомогою методів лінійної оптимізації. Такі методи не зазнають труднощів в обробці локальних екстремумів при навчанні. Однак радіальна базисна функція (РБФ)

нейронні мережі мають один суттєвий недолік: дані мережі володіють поганими екстраполяційними властивостями, також такі мережі дуже громіздкі при великій розмірності вхідного вектору.

Для побудови РБФ нейронних мереж необхідно:

- Наявність еталонних значень вагових векторів нейронів прихованого шару;
- Наявність спеціальної функції активації прихованого шару. Зазвичай використовується функція Гаусса.

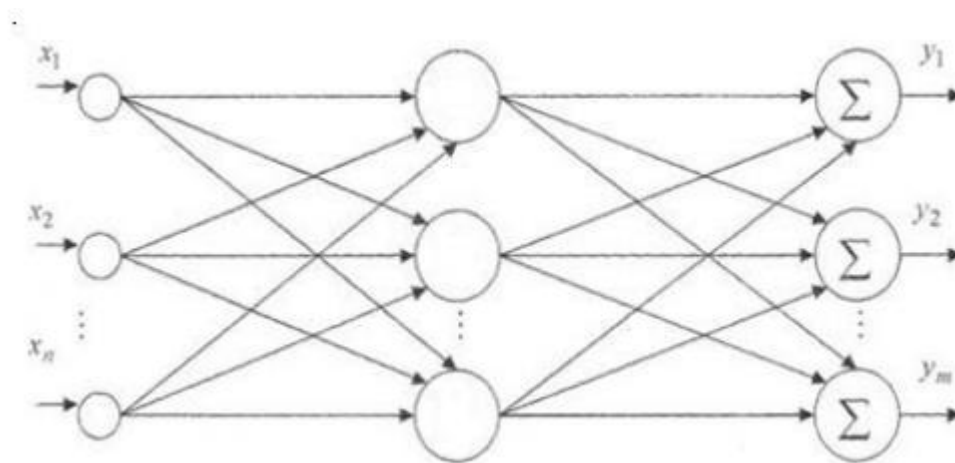


Рис. 2.6 Структурна схема РБФ нейронної мережі

2.3 Мережі зі зворотним зв'язком

Нейронні мережі, що мають з'єднання від виходів до входів, називаються мережами з зворотними зв'язками. У загальному випадку, для мереж із зворотними зв'язками, значення виходів повертаються на входи. А значить, вихід визначається як поточним входом, так і попереднім виходом. Таким чином мережі з зворотними зв'язками мають властивість короткочасної пам'яті. Структурна схема зображена на рис. 2.7 [15].

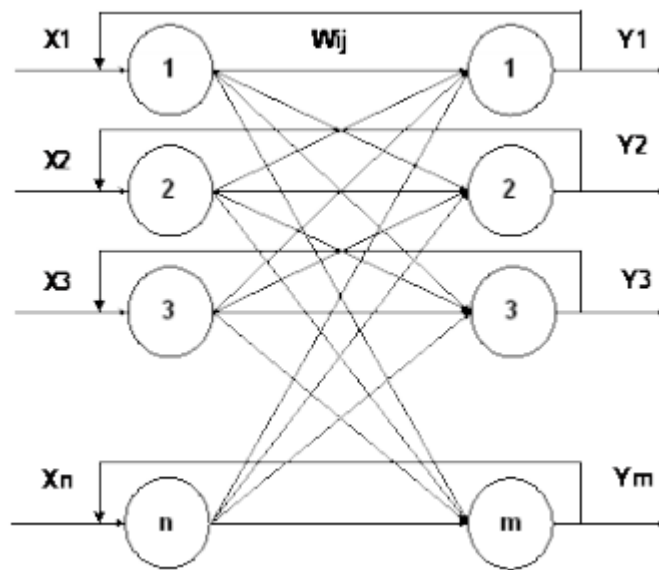


Рис. 2.7 Структурна схема НМ з зворотними зв'язками

2.4 Рекурентні нейронні мережі

Для вирішення завдань прогнозування послідовності значень або порядку проходження об'єкта застосовують рекурентні нейронні мережі (RNN). Такі мережі в процесі роботи можуть зберігати інформацію про своїх попередніх станах.

Найпростішим прикладом такої мережі є мережа Елмана [22]. Рекурентна нейронна мережа складається з трьох шарів: вхідний шар, прихований та вихідний шар; при цьому прихований шар має зворотний зв'язок з самим собою. Принципова схема RNN наведена на рис. 2.8.

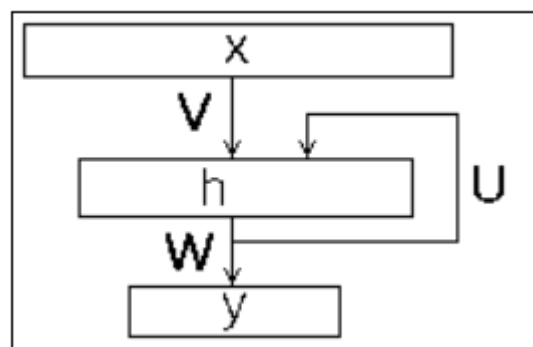


Рис. 2.8 Принципова схема RNN мережі Елмана

На відміну від звичайної мережі прямого поширення, вхідний образ рекурентної мережі це не один вектор, а послідовність векторів. Вектори вхідного образу в заданому порядку подаються на вхід, при цьому новий стан прихованого шару залежить від його попередніх станів. Повторювальний модуль стандартної RNN мережі містить один шар з функцією активації гіперболічного тангенсу.

Основною проблемою таких мереж - проблема довгострокових залежностей. В основі своїй RNN представляють послідовно з'єднані копії однієї і тієї ж нейронної мережі, кожна з яких передає інформацію наступній. Через таку особливість, передана залежність нейрона може втрачатись зі збільшенням відстані між нейронами, здатність використовувати раніше записану інформацію втрачається.

2.5 Довга короткострокова пам'ять (LSTM мережа)

LSTM (англ. Long short-term memory) – різновид рекурентних НМ, який позбавлений проблеми втрати довгострокових зв'язків [22]. LSTM мережі складаються з аналогічного повторювального модуля, однак замість одного, така мережа містить в собі чотири шари, які безпосередньо взаємодіють між собою.

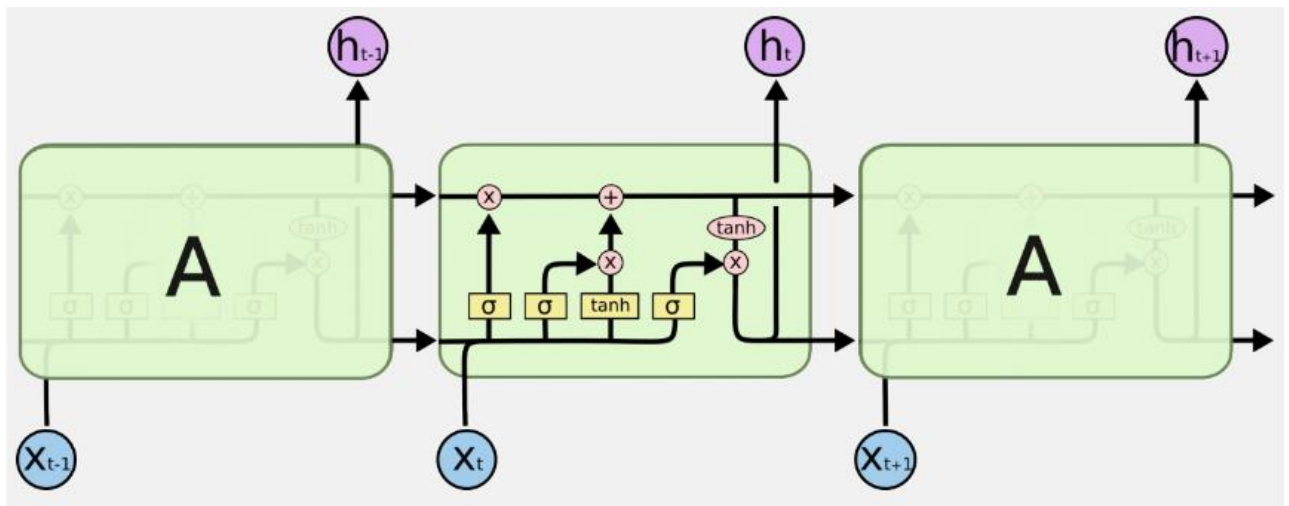


Рис. 2.9 Повторювальний модуль LSTM мережі

Ключовою особливістю LSTM-мережі є стан комірки, представлене горизонтальною лінією у верхній частині наступного р

ис. 2.9. Стан комірки проходить усі операції поступово лише з незначними лінійними взаємодіями.

LSTM-мережа має можливість видаляти і додавати інформацію в стан комірки. Цей процес регулюється спеціальними структурами, званими Гейт.

Гейт - це механізм, що дозволяє пропускати інформацію вибірково. Він складається з шару, функція активації якого – сигмоїда, і операції поточечного множення. Виходом такого шару є 0 чи 1, де при нулі гейт не пропускає інформацію, при одиниці пропускає всю інформацію.

На першому етапі НМ вирішує яку інформацію треба видалити, рішення приймає сигмоїдальний шар, який дає на виході число від 0 до 1. Такий шар називається гейтом забування. Наступним етапом НМ вирішує яку інформацію потрібно записати у стан комірки. Умовно цей етап можна поділити на дві частини: вхідний гейт – за допомогою сигмоїдальної функції вирішується які значення треба оновити. Потім шар з гіперболічним тангенсом у якості функції активації створює новий вектор значень, які можуть бути додані у стан комірки.

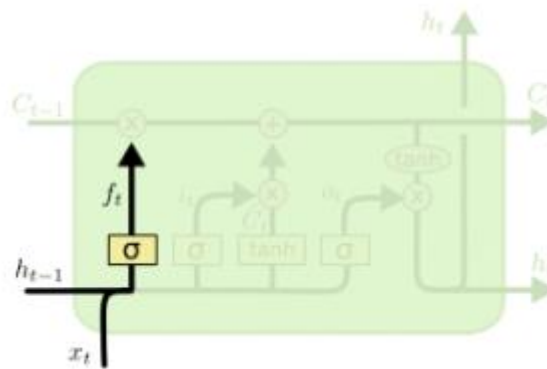


Рис. 2.10 Схема гейту забування

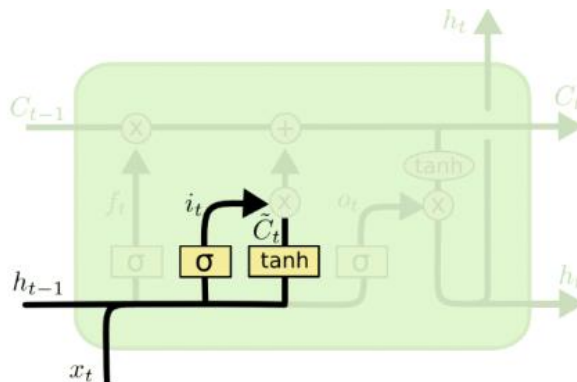


Рис. 2.11 Схема гейту запису

Останнім кроком є рішення НМ о вихідних значеннях. Вихід буде являти собою відфільтрований стан комірки. Спочатку сигмоїдальний шар вирішує, які елементи стану комірки необхідно передати на вихід. Потім стан комірки

перетворюється за допомогою \tanh -шару до інтервалу від -1 до 1 і множиться на вихід сигмоїдального шару, щоб вивести тільки те, що було вирішено вивести.

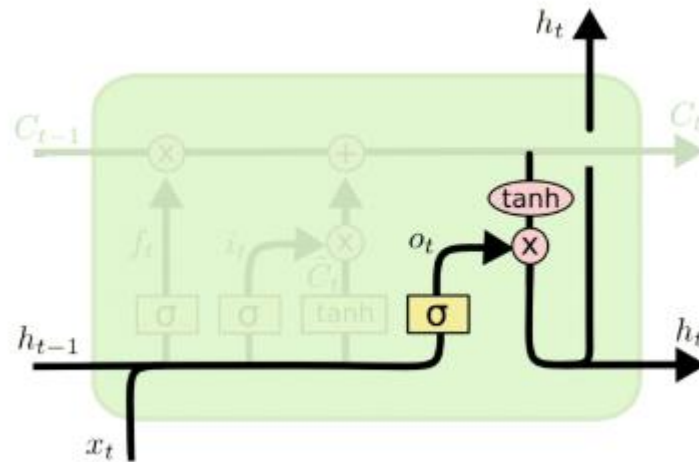


Рис. 2.11 Вихідний шар LSTM мережі

2.6 Методи і алгоритми навчання нейронних мереж

Після вибору типу і архітектури нейронної мережі, щоб застосовувати НМ для рішення прикладних задач, її слід навчити. Мережа навчається для деякого безлічі входів давати бажане безліч виходів. Навчання відбувається за рахунок послідовного пред'явлення вхідних даних з одночасним корегуванням ваг відповідно до певної процедури. Після навчання, кожен вектор, поданий на вхід мережі виробляє вихідний.

Існує два основних методи навчання: з вчителем або без вчителя. Навчання з вчителем передбачає, що для кожного вхідного вектору існує цільовий вектор - цільовий вихід. Вхідний вектор і цільовий вихід прийнято називати навчальною парою [16]. Зазвичай мережа навчається на кількох навчальних парах. Мережа обчислює вихідний вектор, який порівнюється з цільовим вектором, різниця вихідного та цільового векторів за допомогою зворотного зв'язку подається на мережу, ваги змінюються відповідно до алгоритму мінімізації помилки. Вхідні вектори подаються послідовно, обчислюються помилки, ваги підлаштовуються

для кожного вектору до тих пір, поки помилка не досягне допустимого рівня по всьому навчальному масиву.

Більшість сучасних методів навчання без вчителя побудовані на концепції Хеба. Відповідно до цієї моделі навчання НМ, вагові коефіцієнти зростають, якщо активовано обидва нейрона: вхідного і вихідного нейрона. У ШНМ ваги визначаються як добуток рівнів збудження передавального і приймаючої нейрона [16].

$$w_{ij}(n + 1) = w(n) + \alpha NET_i NET_j \quad (2.4)$$

Де $w_{ij}(n)$ – значення ваги нейрона від i до нейрона j до корегування, $w_{ij}(n + 1)$ – значення ваг від i до нейрона j після корегування, α – коефіцієнт швидкості навчання, NET_i – вихід нейрона i , NET_j – вихід нейрона j .

Різниця між бажаним часовим рядом і рядом, отриманим в результаті моделювання називається помилкою навчання. Навчання проводиться для мінімізації помилки шляхом ручної або автоматичної корекції ваг зв'язків. Помилка для конкретного типу мережі визначається шляхом обчислення різниці між цільовими і вхідними значеннями для всіх наявних спостережень. Сформована площа називається функцією помилок або критерієм якості навчання.

Для лінійних функцій активації нейронів є можливість побудови алгоритму досягнення абсолютного мінімуму помилки навчання. У свою чергу, для нелінійних функцій активації неможливо гарантувати досягнення глобального мінімуму.

Для такого підходу навчання слід виділити геометричний аналіз поверхні помилок [16]. При такому підході визначається загальне число N ваг і зміщень як вільні параметри моделі. Для кожного набору параметрів визначається відповідне вимір у вигляді помилки. Для кожного поєднання ваг і зміщень відповідна помилка мережі зображується на $N + 1$ мірному просторі. Така поверхня називається поверхнею функції помилок. При застосуванні такого

підходу навчання проводиться до пошуку максимального допустимого значення мінімуму на поверхні.

Лінійні моделі мережі мають функцію помилок у вигляді суми квадратів - така поверхня являє параболоїд з єдиним мінімумом. Поверхня помилок нелінійної моделі може мати кілька локальних мінімумів, плоскі ділянки і т.д. Визначення глобального мінімуму для багатовимірних моделей аналітично неможливо.

Також слід виділити алгоритм зворотного поширення помилки - алгоритм навчання нейронних мереж [16]. При використанні даного алгоритму, НМ обчислює помилку на вихідному шарі і розраховує вектор градієнта ваг і зміщень. Отриманий вектор вказує напрямок найкоротшого спуску по поверхні для даної точки. Отже, по мірі просування по вектору градієнта, помилка зменшується. Для такого методу важливим параметром є визначення частоти дискретизації досліджуваного часового ряду.

При навчанні нейронної мережі важливою вимогою є мінімізація помилки при появі в мережі нових даних. Випадок, коли мінімізується не та помилка, яка необхідна, називається явищем перенавчання [16]. Кожен набір даних можна представити у вигляді апроксимації деякої функції многочленом. Ступінь многочлена підвищується в залежності від складності форми розглянутого часового ряду. Для кожного набору даних можна підібрати апроксимуючий многочлен для отримання максимально правдоподібною математичної моделі. В такому випадку, многочлен низького порядку може виявитися досить грубим для апроксимації даних, в той час як многочлен високого порядку може точно слідувати даним, що не має ніякого відношення до форми істинної залежності.

Отже, мережі з великою кількістю ваг дозволяють відтворювати дуже складні функції, і в цьому сенсі вони схильні до перенавчання. Мережа ж з невеликою кількістю ваг може виявитися недостатньо гнучкою, щоб змодельовати наявну залежність.

Для запобігання ефекту перенавчання використовується механізм контрольної перевірки [16]. Частина загальних даних зберігається і не використовується при навчанні та застосовується для незалежного контролю. При навчанні мережі помилка на контрольній множині буде спадати до тих пір, поки убуває помилка при навчанні. Якщо ж контрольна помилка перестала зменшуватися або стала рости, значить НМ стала занадто швидко слідувати вихідними даними, а, отже, процес навчання слід зупинити.

ВИСНОВОК ДО РОЗДІЛУ 2

Підсумовуючи вищесказане, для вирішення завдання прогнозування часових рядів необхідно визначити Найбільш ефективну конфігурацію нейронної мережі з урахуванням особливостей даного ряду. Важливим фактором правильності налаштування мережі є процес навчання і подальшого її корегування, а також виявлення ефекту перенавчання. Конфігурація нейронної мережі може задаватися багатьма факторами, наприклад: кількість входів і виходів, кількість проміжних шарів, функція активації, зв'язок між шарами.

Ефективність роботи НМ оцінюється за допомогою обчислення середньоквадратичне відхилення (СКВ) або абсолютної помилки. А значить, завдання прогнозування часових рядів можна описати у вигляді оптимізації багато параметричної функції.

А отже, можна виділити три етапи побудови нейронної мережі:

- Вибір початкової конфігурації: кількість шарів, число нейронів, кількість вхідних даних;
- Моделювання та навчання НМ з оцінкою якості навчання або з оцінкою контрольної помилки;
- Виявлення ефекту перенавчання та коригування нейронної мережі.

Підсумувавши вищесказане, можна скласти методику побудови математичної моделі:

- Складання якісної моделі, тобто виявлення законів зв'язку, головних і характерних рис. досліджуваного процесу. Для фізичних процесів - формулювання фізичних закономірностей процесу.
- Створення математичної моделі:
 1. Детермінований процес або стохастичний;
 2. Відокремлення суттєвих факторів: якщо в системі діє кілька факторів одного порядку, то всі вони повинні бути враховані, або відкинуті;
 3. Виділення додаткових умов, наприклад: початкових, граничних, умов сполучення.
- Вивчення математичної моделі:
 1. Математичне обґрунтування моделі - обґрунтування коректності моделі;
 2. Якісне дослідження моделі - з'ясування поведінки моделі в крайніх і граничних ситуаціях;
 3. Чисельне дослідження моделі: розробка алгоритму, розробка чисельних методів дослідження моделі, створення і реалізація програми, комп'ютерний експеримент.
- Отримання результатів і їх інтерпретація. Порівняння початкових даних з результатом аналізу.
- Використання отриманих результатів.

РОЗДІЛ 3

ДОСЛІДЖЕННЯ ДЕТЕРМІНОВАНИХ ПРОЦЕСІВ

Для вирішення завдання прогнозування детермінованих процесів необхідно визначити ефективну конфігурацію нейронної мережі. Вхідні дані представлені в вигляді монотонних процесів і процесу з екстремумом, а також процесу в межах деякого інтервалу часу. У якості монотонних процесів розглянуті наступні: $y(t) = k * t$, $y(t) = 1 - \exp^{-at}$. У якості процесу з екстремумом досліджена наступна ділянка синусоїди: $y(t) = A * \sin(w * t)$, для обох випадків було використано значення $t = 5$ с.

3.1 Лінійна нейрона мережа

Для прогнозування лінійного тренду детермінованих процесів була використана одношарова статична лінійна нейронна мережа. Дана НМ дозволяє без будь-яких додаткових зусиль проаналізувати часовий ряд. Принципова схема даної нейронної мережі представлена на рис. 3.1. Єдина відмінність даної НМ від персептрона полягає в наявності лінійної функції активації, яка дозволяє НМ виявити лінійну залежність в даному часовому ряді. Використання лінійної нейронної мережі обумовлено на припущенні, що вхід (вектор Р) і вихід (вектор Т) пов'язані між собою лінійно.

За допомогою пакета прикладних програм MATLAB, а також додаткового пакета Deep Learning Toolbox була побудована статична одношарова лінійна нейронна мережа з допомогою функції `newlind` [19]. Статична НМ характеризується відсутністю ліній затримки або зворотних зв'язків. У доповненні наведено приклад налаштування лінійної нейронної мережі з використанням функції `newlind`. Схема змодельованої НМ зображена на рис. 3.1.

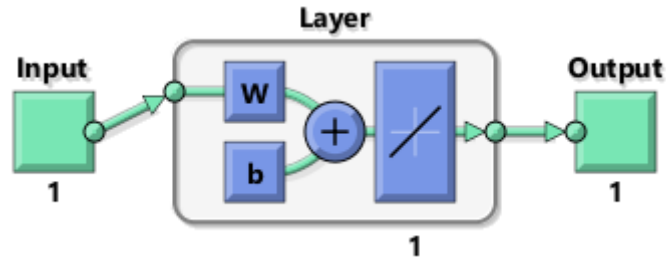


Рис. 3.1 Модель лінійної статичної НМ

Така НМ здатна до лінійної апроксимації часового ряду, що видно з результатів моделювання з монотонними функціями. Результати моделювання для значень $k = 1,5$ і $a = 0,5, 2$ наведені нижче.

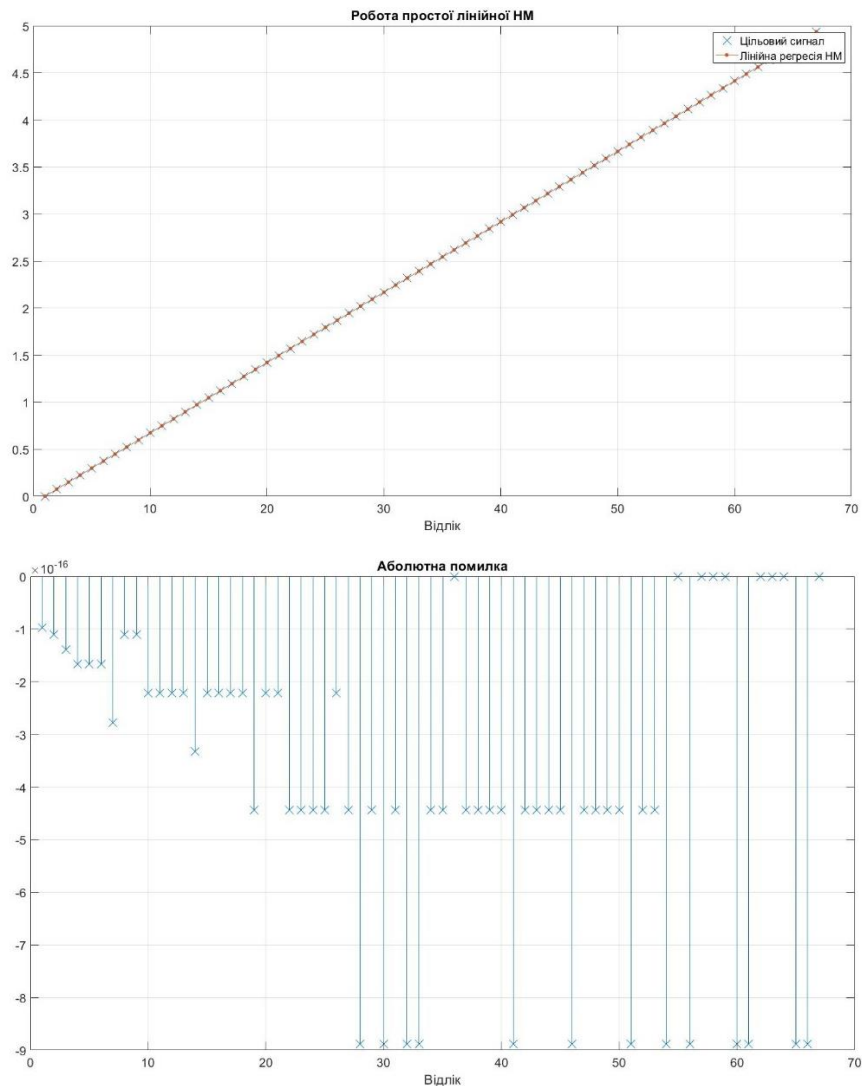


Рис. 3.2 Функція $y(t) = k * t$ для $k = 1$, а – результат роботи НМ, б – розрахунок абсолютної помилки

Результат нейронної мережі для інших значень k не буде суттєво від відрізнятися від значення $k = 1$. Помилка у даному випадку майже нульова. Використання нелінійних функцій активації для визначення лінійної залежності ряду не буде доцільним, так як у такому випадку функція активації буде обмежена у деякому діапазоні, що, у свою чергу, призведе до виникнення великої помилки.

Для іншої монотонної функції було проведено аналогічний експеримент без зміни конфігурації НМ. Як можна побачити з результатів роботи нейронної мережі, лінійна мережа здатна визначати тренд та лінійні зміни у монотонній функції. Розраховані середні квадратичні відхилення збільшуються з збільшенням коефіцієнта a . Результати роботи НМ з монотонною незгасаючою функцією наведені на рис. 3.3 та рис. 3.4 для $a = 0.5$ та $a = 2$ відповідно.

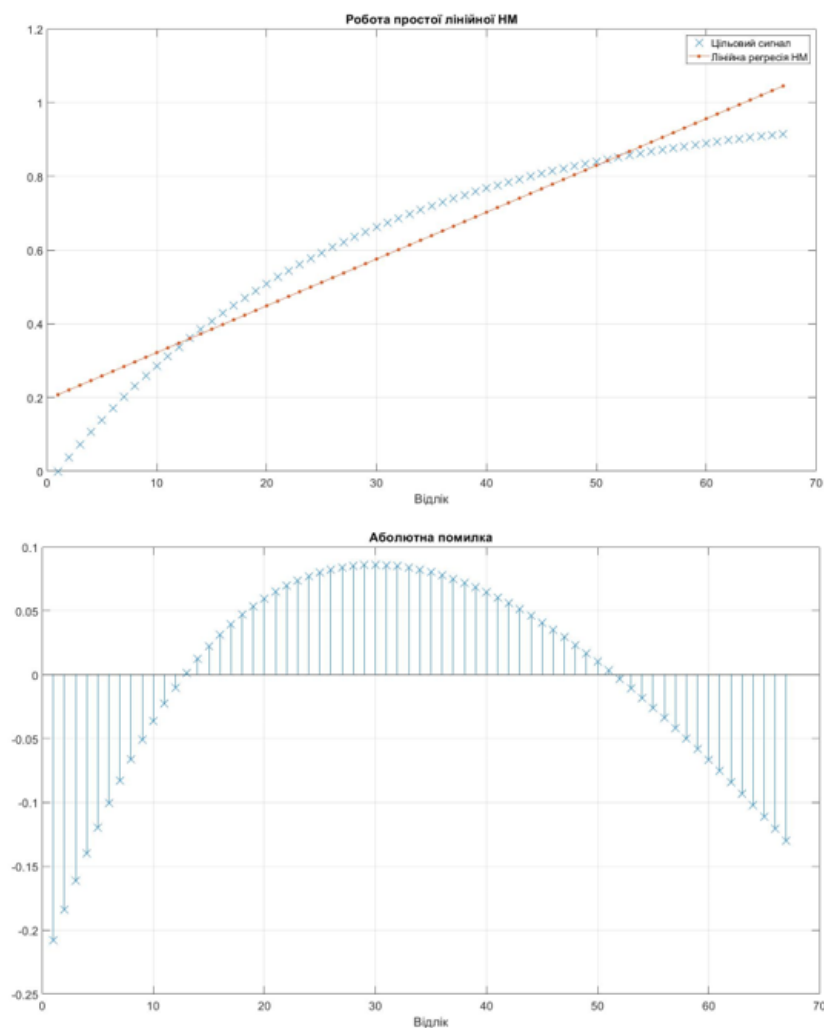


Рис. 3.3 Результати роботи НМ та помилка для $a = 0.5$

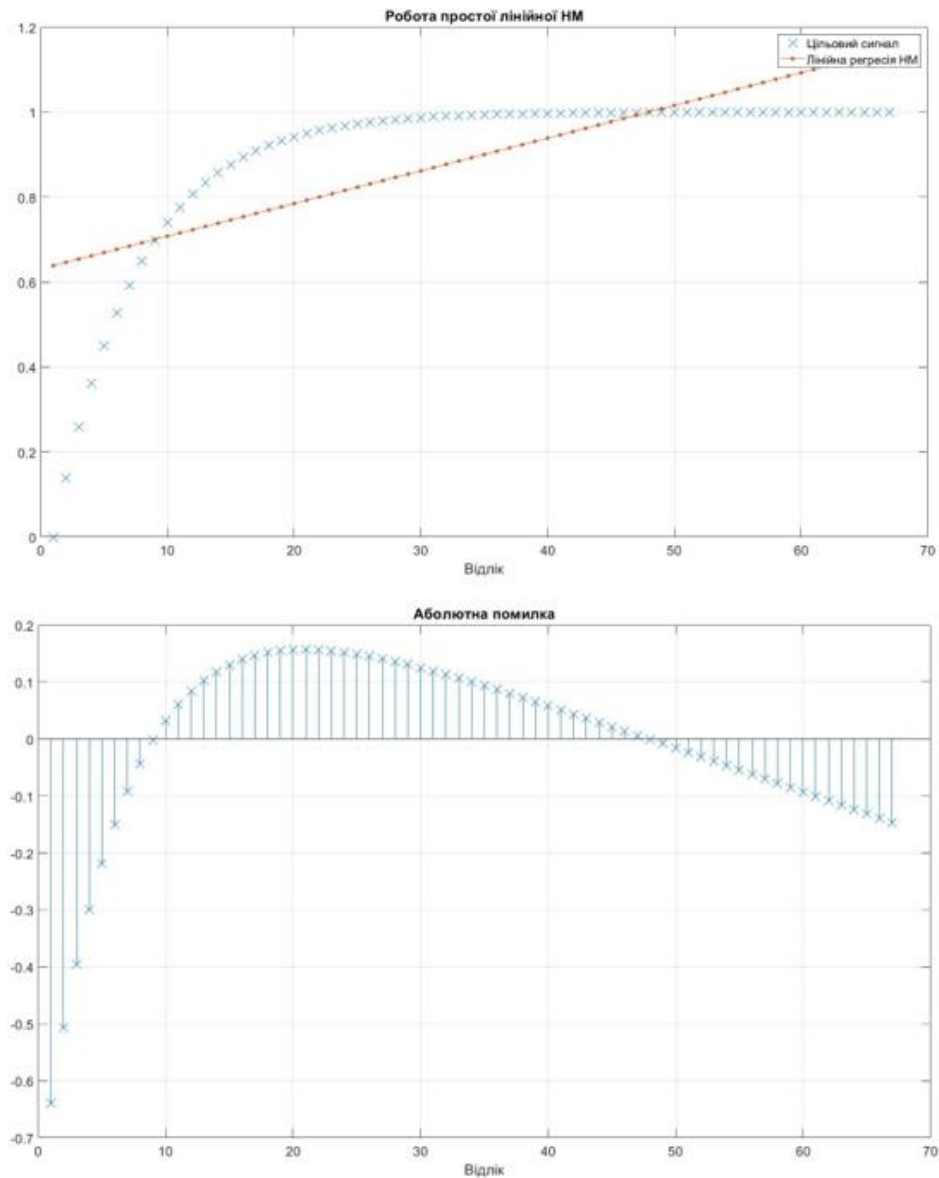


Рис. 3.4 Результати роботи НМ та помилка для $a = 2$

СКВ ($a = 0.5$) = 0.0777, СКВ ($a = 2$) = 0.1550.

Функція $y(t) = A * \sin(w * t)$ містить у собі частини функцій $y(t) = k * t$ та $y(t) = 1 - \exp^{-at}$. Також у цьому процесі спостерігається локальний екстремум, після якого синусоїда змінює знак руху, відрізок після екстремуму спадний. Розглянута частина синусоїди розбита на три однакових відрізків, у якому спостерігаються декілька процесів: монотонні та знакозмінні. Конфігурація синусоїди наступна: $A = 5, w = 0.47, t = 5$. Результат роботи НМ на повному проміжку часового ряду наведений нижче.

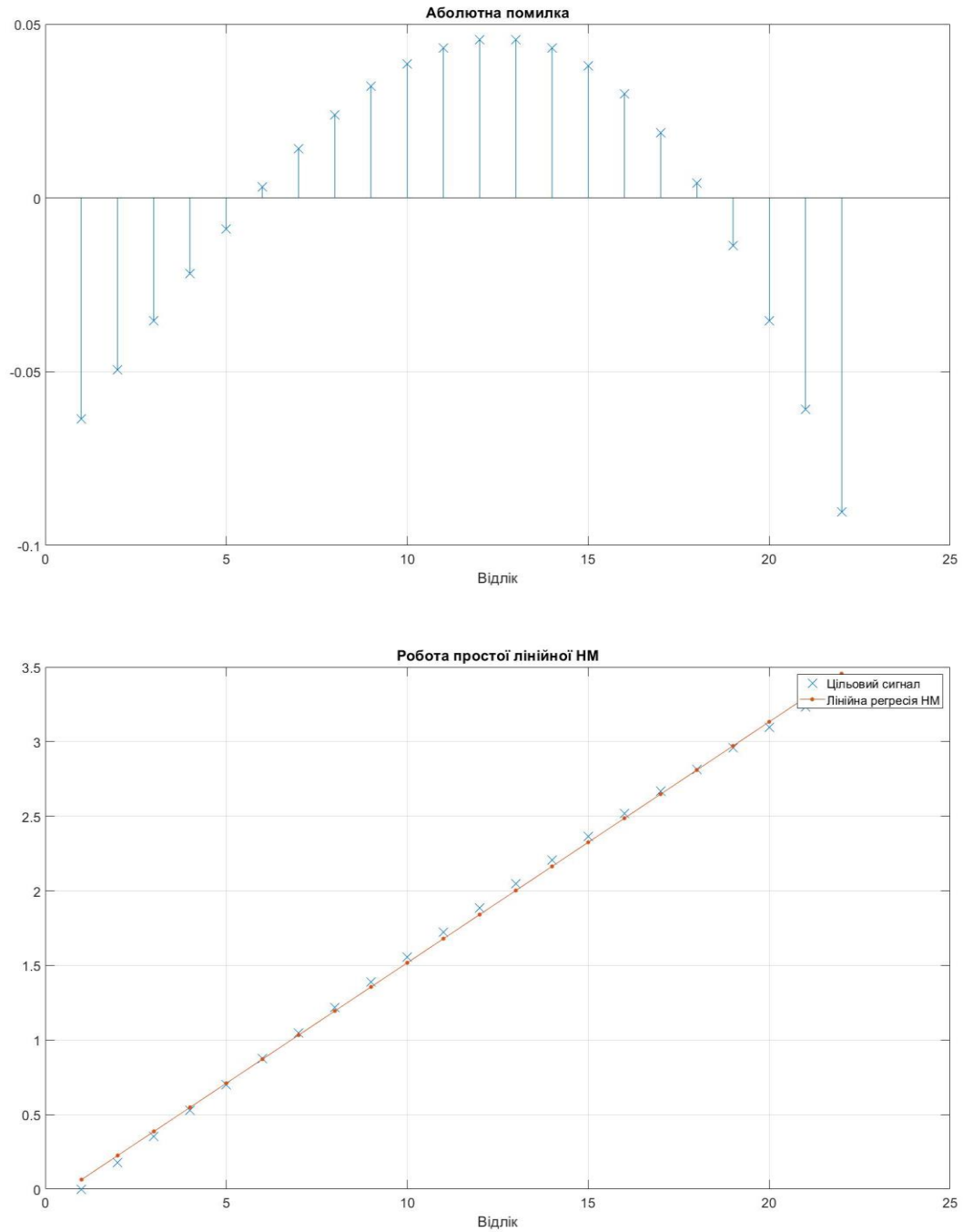


Рис. 3.5 Перша третина синусоїди, а – результат роботи НМ, б – абсолютна помилка

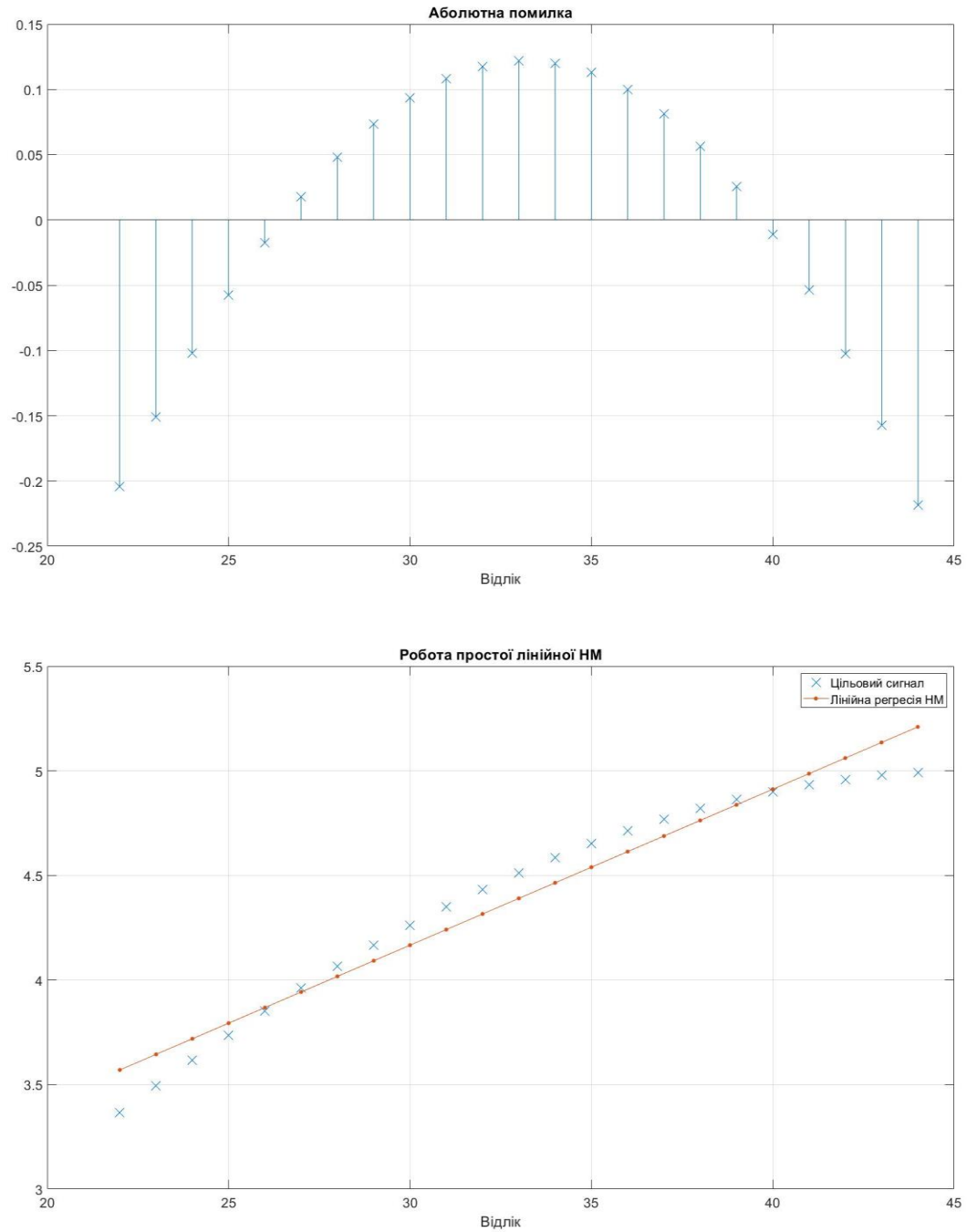


Рис. 3.6 Друга третина синусоїди, а – результат роботи НМ, б – абсолютна помилка

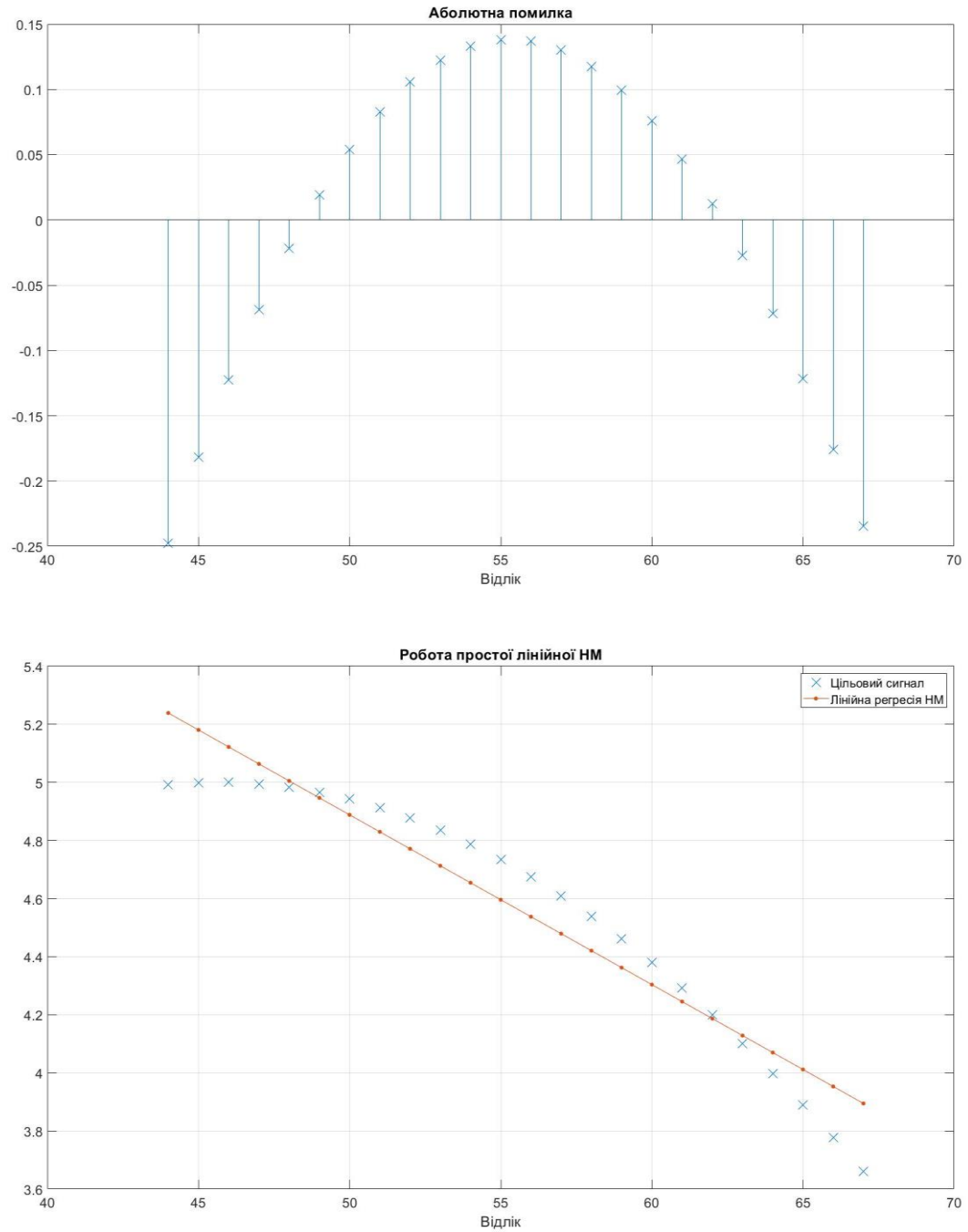


Рис. 3.7 Остання третина синусоїди, а – результат роботи НМ, б – абсолютна помилка

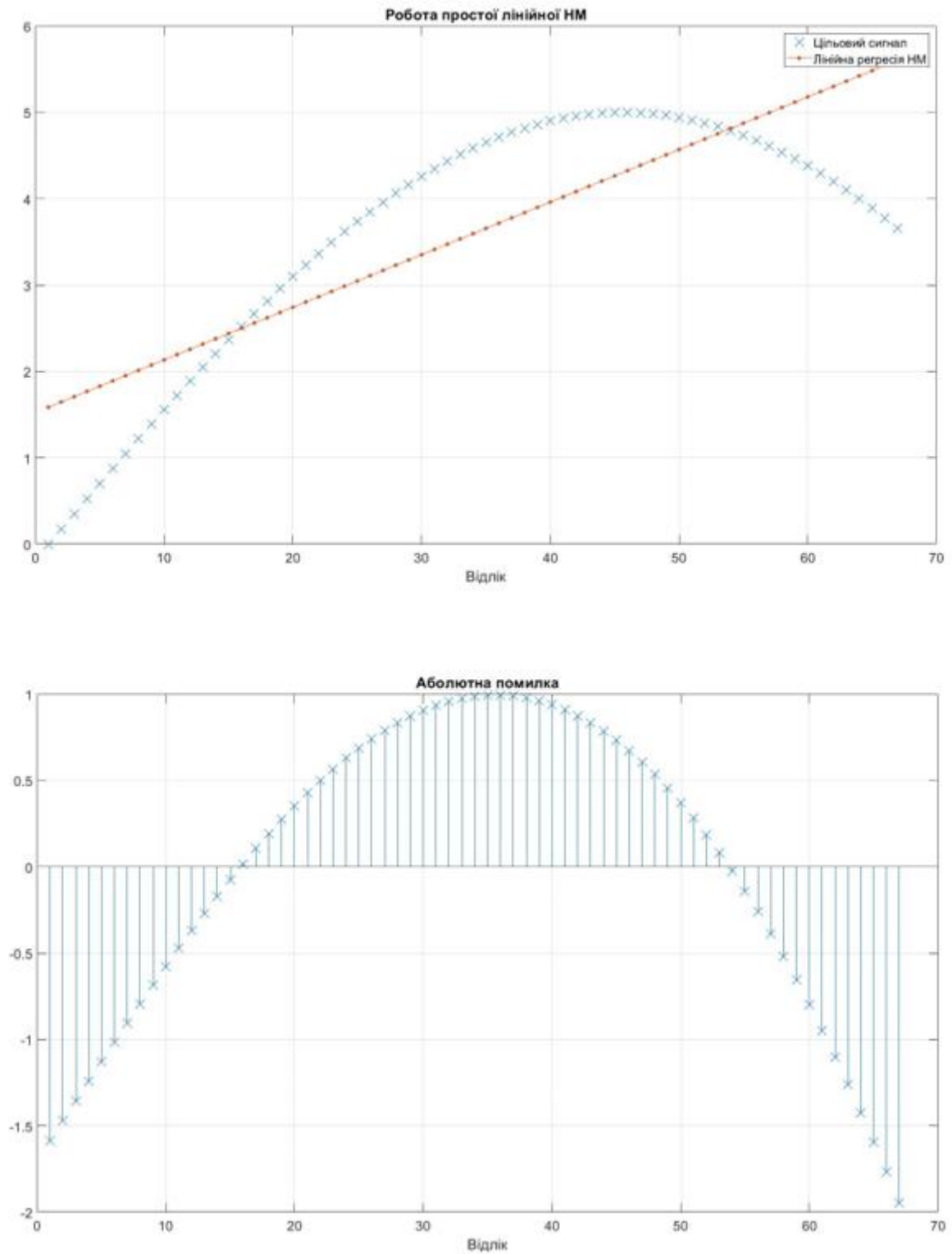


Рис. 3.8 Функція $y(t) = A * \sin(w * t)$, а – результат роботи НМ, б – абсолютна помилка

З наведених вище результатів, можна побачити, що статична лінійна НМ здатна визначати тренд процесу на всьому проміжку вхідних даних. Але така нейронна мережа не здатна прогнозувати наперед.

Апроксимація за допомогою статичної лінійної НМ дозволяє досліджувати числові характеристики і якісні властивості об'єкта, зводячи завдання до вивчення більш простих або більш зручних об'єктів, наприклад, тих, параметри яких легко обчислюються або відомі заздалегідь.

Лінійна регресія була першим видом регресійного аналізу, який був ретельно вивчений і почав широко використовуватися в практичних додатках. Це пов'язано з тим, що в лінійних моделях оцінювання параметрів простіше, а також з тим, що статистичні властивості отриманих оцінок легше визначити.

Лінійна регресія має багато практичних застосувань. Більшість додатків потрапляють в одну з двох широких категорій:

- Якщо метою є прогнозування, лінійну регресію можна використовувати для підгонки моделі до спостережуваного набору даних.
- Якщо мета полягає в тому, щоб пояснити змінність вихідної змінної, можна застосувати лінійний регресійний аналіз для кількісної оцінки сили взаємозв'язку між вихідними і вхідними змінними.

3.2 Рекурентна нейронна мережа LSTM

Задля чисельного прогнозування було використано рекурентну НМ та функцію $y(t) = A * \sin(w * t)$, $A = 5, w = 0.47, t = 5$. Максимальна частота гармонійної складової $f_{max} = \frac{w}{2\pi} = 0.074$, період дискретизації не має перевищувати $T_d \leq \frac{1}{2 * f_{max}}$. Обраний період дискретизації $T_d = 0.075$.

У якості рекурентної нейронної мережі було взято LSTM - різновид архітектури рекурентних нейронних мереж [20].

Для прогнозування значень майбутніх кроків часового ряду мереже LSTM регресії була натренована методом sequence-to-sequence [21], де вихід - це навчальні послідовності зі значеннями, які зміщені на один крок часу. Тобто на кожному етапі послідовності входу, мережа LSTM вчиться передбачати значення наступного кроку часу.

Для прогнозування значень кількох часових кроків було використано функцію predictAndUpdateState для прогнозування часових кроків один за одним та оновлення стану мережі при кожному передбаченні.

Вхідні дані були розділені таким чином, щоб дані, які будуть використовуватися для тренування охоплювали локальний екстремум. Дані розділені наступним чином: 80% для тренування, 20% для тестування. А отже, для аналізу було взято часовий ряд з 67 точками, де 53 точки відокремлені для тренування, останні 14 - для прогнозування. Для кращої адаптації НМ та для запобігання розбіжностей тренувань вхідні дані були пронормовані.

Виходячи з отриманого часового ряду, для першого налаштування НМ було використано 200 ітерацій навчання, встановлений поріг градієнта, який дорівнює одиниці. Початкова швидкість навчання дорівнює 0.005, а після 125 ітерацій швидкість навчання буде помножена на 0.2.

За методом тренування НМ sequence-to-sequence прогнозування буде вестись наступним чином: нейронна мережа прогнозує наступне значення часового ряду, починаючи з початку, порівнює отримане значення з цільовими вектором та корегує вагові коефіцієнти. Після тренування та корекції ваг НМ, мережа починає прогнозувати невідоме значення часового ряду.

Після налаштування НМ, було здобуто перший результат роботи з довільними значеннями прихованих шарів та максимальної кількості ітерацій навчання. Результат тренування та роботи нейронної мережі наведені на рис. 3.9

та рис. 3.10, рис. 3.11 відповідно. Слід зазначити, що кількість ітерацій навчання були підібрані довільним шляхом.

З рис. 3.9 можна побачити, що після 150 ітерації для заданої кількості шарів зміна середньоквадратичного відхилення майже не відбувається. А отже, для подальшого налаштування даної НМ мінімальною кількістю ітерацій навчання буде дорівнювати 150.

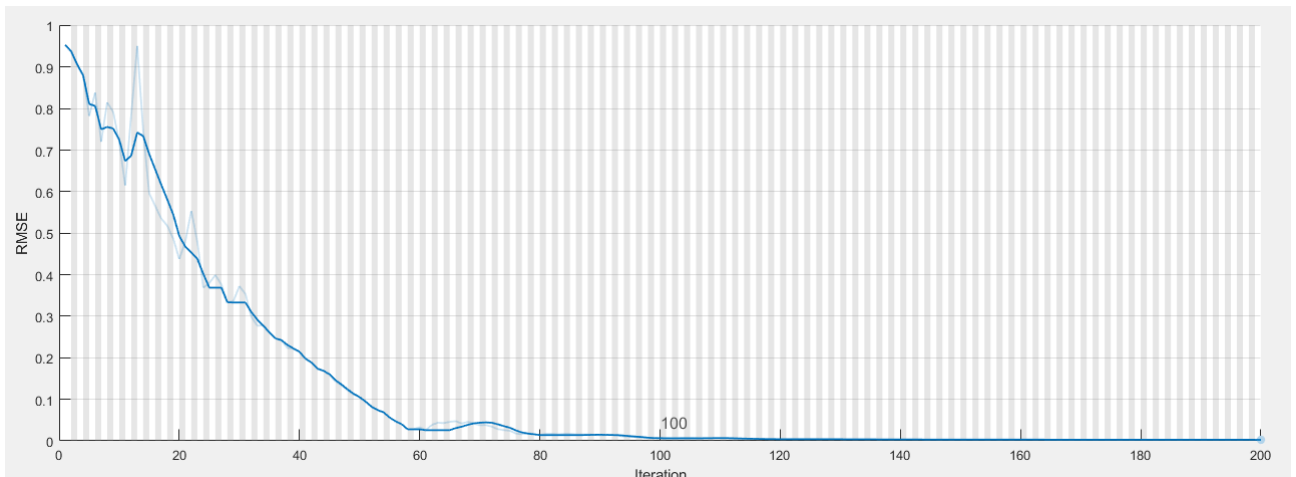


Рис. 3.9 Результат тренування LSTM нейронної мережі

Як можна побачити з рис. 3.10 та рис. 3.11, результат роботи рекурентної мережі, яка містить 150 прихованих шарів та 200 ітерацій навчання, можна вважати майже задовільним.

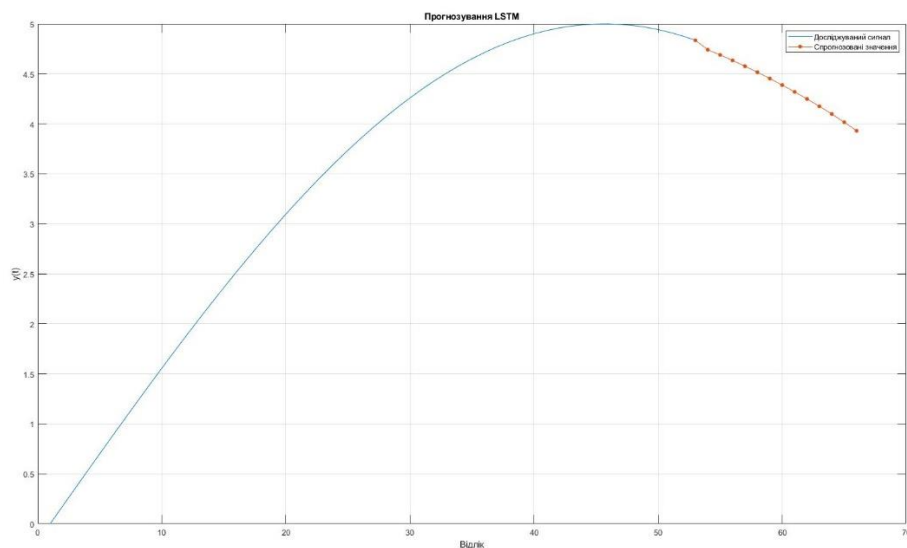


Рис. 3.10 Результат тестування LSTM мережі

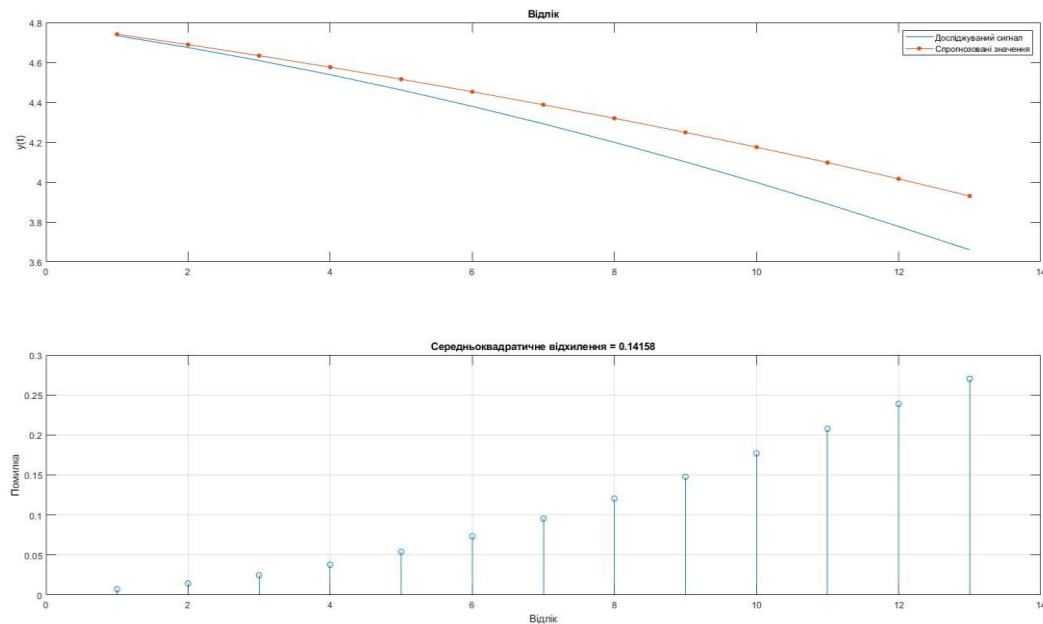


Рис. 3.11 Помилки при тестуванні LSTM мережі, СКВ = 0.14

Задача більш точного прогнозування за допомогою рекурентної нейронної мережі була вирішена тестовим методом пошуку кількості шарів НМ та кількості ітерацій навчання. Даний тип НМ характеризується

Для правильної та ефективної роботи НМ необхідно підібрати оптимальне значення кількості ітерацій тренувань для заданої кількості прихованих шарів. Для кожного тестування було вираховано середньоквадратичне відхилення. Для подальшого дослідження НМ була відібрана нейрона мережа з найменшим СКВ.

Як можна побачити з рис. 3.12, при кількості прихованих шарів 175 та кількості ітерацій навчання 185 нейрона мережа показала найкращий показник середньоквадратичного відхилення, який дорівнює 0.0138.

Результат роботи оновленої натренованої нейронної мережі наведено на рис. 3.13 та рис. 3.14

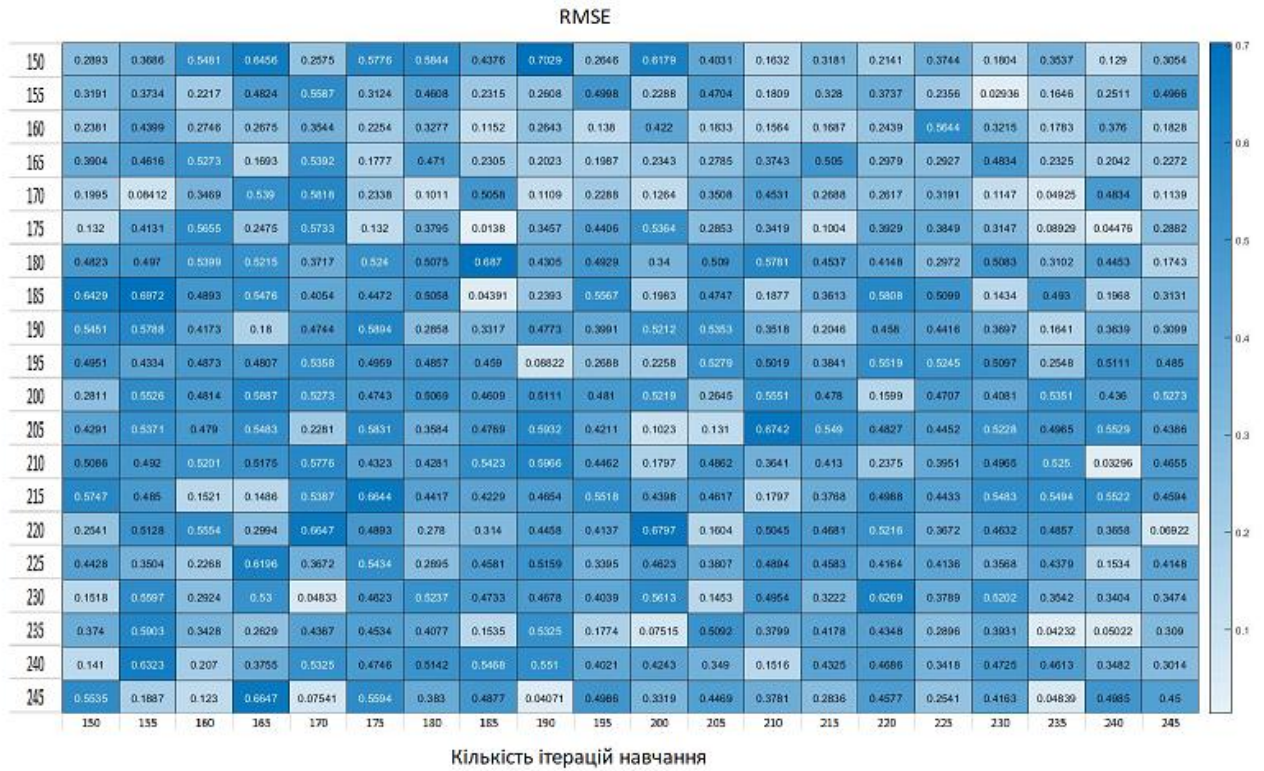


Рис. 3.12 Показник СКВ для різних значень прихованих шарів та ітерацій навчання

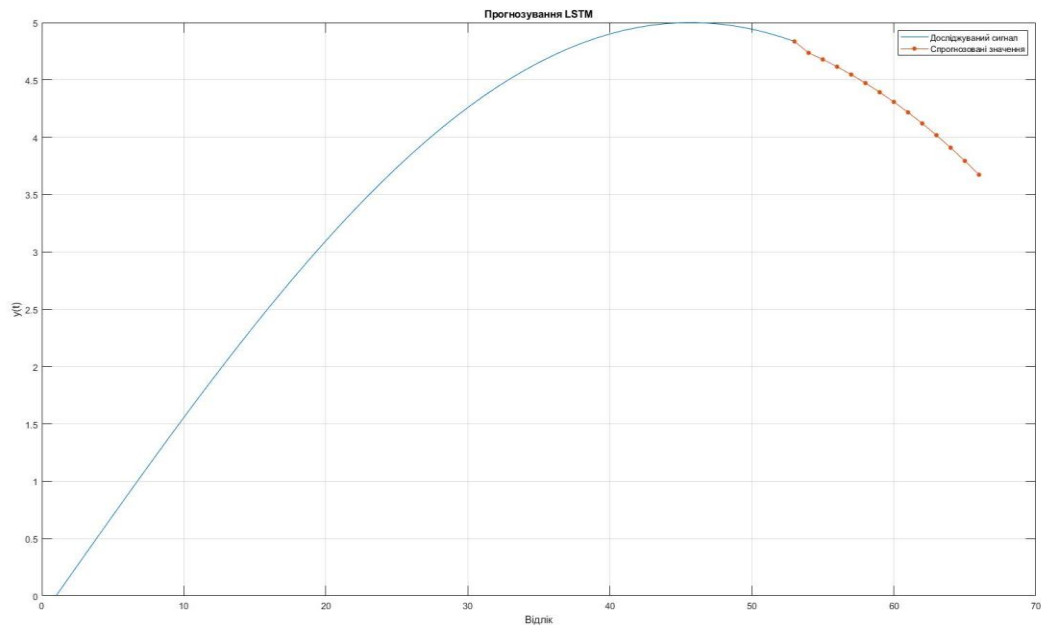


Рис. 3.13 Результат роботи натренованої LSTM нейронної мережі

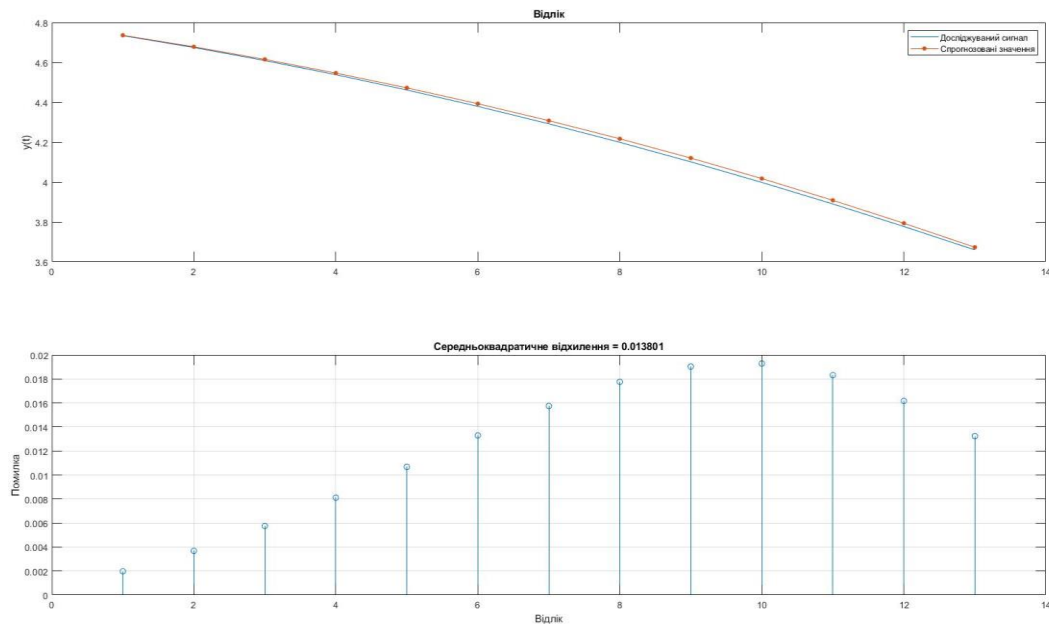


Рис. 3.14 Помилка роботи натренованої LSTM нейронної мережі

З рис. 3.13 можна побачити, що прогнозовані значення даної рекурентної НМ майже повторює графік синусоїди, максимальне відхилення від цільових значень досягає при прогнозуванні десятого відліку тестованої вибірки. Розраховане СКВ для цієї вибірки складає 0.0138, що у свою чергу свідчить про якісне налаштування НМ та її здатність до прогнозування часових рядів.

Результат роботи рекурентної залежить в першу чергу від кількості та інформативності вхідних даних. Така НМ не здатна належним чином прогнозувати наступні значення, якщо розглянутий часовий ряд пропускає або не охоплює локальні екстремуми. При належному навчанні рекурентні НМ здатні до більш точного прогнозування. Також для таких НМ властиво явище перенавчання, яке виправляється зменшенням кількості ітерацій навчання. А кількість прихованих шарів визначається таким чином, щоб СКВ при тренуванні максимально надходило до 0 при заданій кількості ітерацій тренувань.

ВИСНОВОК ДО РОЗДІЛУ 3

Розглянуті нейронні мережі суттєво відрізняються одна від одної, як за архітектурою й принципом роботи, так і за отриманим результатом. Лінійна статична нейронна мережа підходить для аналізу тенденцій, які відбувались у процесі у певні проміжки часу. Така мережа проста у використанні і здатна швидко визначити лінійні залежності у досліджуваному процесі. Великі абсолютні помилки обумовлені наявністю нелінійності процесу, в свою чергу для лінійних процесів, така НМ показує гарні результати роботи. Але для аналізу часових рядів, які припускають наявність нелінійності у процесі, лінійні НМ не підходять, тому в таких випадках слід звертатися до інших типів мереж. У якості такої мережі було використано рекурентну LSTM нейронну мережу. Дана НМ здатна впоратись з прогнозуванням майбутніх значень числового ряду, що, в свою чергу, можна побачити з результатів роботи нейронної мережі: з результату тесту, СКВ натренованої мережі складає 0.013801, а відхилення від цільового значення часового ряду не перевищує 0.02.

Підсумовуючи результати дослідження, видно, що статична лінійна нейронна мережа, побудована на базі архітектури перспетрона може використовуватися задля визначення тренду всього часового ряду, що розглядається. Але така НМ не може впоратись з чисельним прогнозуванням майбутніх значень.

Задля чисельного прогнозування слід використовувати інші архітектури нейронних мереж, одна з таких – рекурентна мережа LSTM. Проте такі мережі складніші для аналізу та налаштуванні. LSTM мережі краще себе показують при використанні довгого часового ряду, який повинен містити локальні екстремуми досліджуваного процесу.

ВИСНОВКИ

Метою даної дипломної роботи було дослідження можливостей нейронних мереж в прогнозуванні часових рядів, будування методичного та алгоритмічного забезпечення для прогнозування фізичних процесів на основі нейромережевих технологій.

У першому розділі було досліджено проблеми, які виникають під час аналізу та прогнозування показників різних галузей, а також задачі, що мають вирішуватись під час дослідження. Продемонстровані основні варіанти вирішення завдань і проблем прогнозування. Описані основні проблеми стандартних статистичних методів прогнозування. Проведено порівняння різних алгоритмів прогнозування. У другому розділі оглянуто основні конфігурації нейронних мереж. Проаналізовані основні етапи реалізації НМ в задачах прогнозування, а також підсумовані етапи побудови математичної моделі.

Третій розділ присвячений дослідженню деяких нейронних мереж у задачах прогнозування. Для аналізу були використані статична лінійна та рекурентна мережі. Аналіз проводився на детермінованих процесах, а саме: монотонні та з екстремумом. Було проаналізовано основні проблеми, які вирішують наведені НМ, їх недоліки. Було побудовано два типи нейронних мереж, а також розроблено відповідне програмне забезпечення за допомогою пакету прикладних програм Matlab та додатку Deep Learning Toolbox, який дозволяє скласти архітектуру наведених нейронних мереж. Проведено аналіз та навчання рекурентної нейронної мережі LSTM. Отримані результати дозволяють стверджувати про достатню точність та можливість використання даних мереж для прогнозування фізичних процесів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- 1 Математичне моделювання систем і процесів [Текст]: Навч. посіб. / Павленко П. М., Філоненко С. Ф., Чередніков О. М., Трейтяк В. В. – НАУ, 2017. – 23 с.
- 2 Методика та організація наукових досліджень [Текст]: Навч. посіб. / С. Е. Важинський, Т. І. Щербак. – Суми: СумДПУ імені А. С. Макаренка, 2016. – 97 с.
- 3 Томашевський В. М. Моделювання систем [Текст]. - К: Видавнича група ВНУ, 2005. – 25 с.
- 4 Efficient Market Hypothesis: Is The Stock Market Efficient? [Електронний ресурс] - Режим доступу: <https://www.investopedia.com/articles/basics/04/022004.asp>
- 5 Джек Д. Швагер Технический анализ: Полный курс [Текст]: пер. з англ. А. Куницын, Б. Зуев «Альпина Паблишер», 2001. – 13 с.
- 6 Джон Дж. Мэрфи Технический анализ финансовых рынков. Полный справочник по методам и практике трейдинга [Текст]: пер. з англ. "ИД «Вильямс»", — 2015 13 с.
- 7 Фундаментальный анализ на рынке ценных бумаг [Текст]: Навч. посіб. / Е.В. Семенкова, Н.Ю. Мазаев, О.В. Помаскина А.А. Эдилбаев, Д.Г. Перепелица, Ю.Ю. Финогенова. – ООО «РУСАЙНС», 2017. – 7 с.
- 8 Механізми функціонування нової моделі ринку електричної енергії України [Електронний ресурс] - Режим доступу: <http://reform.energy/media/116/910dcc41916ac055e22a098d68543ee8.pdf>
- 9 Electricity price forecasting: A review of the state-of-the-art with a look into the future [Електронний ресурс] - Режим доступу: <https://www.sciencedirect.com/science/article/pii/S0169207014001083>
- 10 Основы економічного прогнозування [Текст]: Навч. посіб. / Грабовецкий Б.Є. - Вінниця: ВФ ТАНГ, 2000. – 9 с.
- 11 Дубина И. Н. Основы математического моделирования социально-экономических процессов: Навч. посіб. – Юрайт, 2016. – 308 с.

- 12 Методы прогнозирования и модели распространения заболеваний / М. А. Кондратьев // Компьютерные исследования и моделирование 2013 т. 5 № 5 с. 863–882
- 13 Бідюк П.І. Аналіз часових рядів [Текст]: Навч. Посіб. / Бідюк П.І., Романенко В.Д., Тимошук О.Л. – Київ: Політехніка, 2010. – 7 с.
- 14 Прогнозирование временных рядов [Электронный ресурс] - Режим доступа: http://edu.mmcs.sfedu.ru/pluginfile.php/24419/mod_resource/content/1/14%20%D0%9F%D1%80%D0%BE%D0%B3%D0%BD%D0%BE%D0%B7%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5%20%D0%B2%D1%80%D0%B5%D0%BC%D0%B5%D0%BD%D0%BD%D1%8B%D1%85%20%D1%80%D1%8F%D0%B4%D0%BE%D0%B2.pdf
- 15 Уоссермен Ф. Нейрокомпьютерная техника [Текст]/ Ф.Уоссермен. – М.: Мир, 1992. – 180 с.
- 16 Нейронные сети. Реализация в Matlab [Текст]/: Навч. Посіб. /С.Г. Николаева. – Казань: Казан. гос. энерг. ун-т, 2015. – 92 с
- 17 Алгоритми навчання НМ [Електронний ресурс] - Режим доступа: <https://neurohive.io/ru/osnovy-data-science/lstm-nejronnaja-set/>
- 18 Лінійна НМ реалізація в MatLab [Електронний ресурс] - Режим доступа: <https://www.mathworks.com/help/deeplearning/ref/newlind.html>
- 19 Long Short-Term Memory (LSTM) [Електронний ресурс] - Режим доступа: <https://www.mathworks.com/discovery/lstm.html>
- 20 Understanding Encoder-Decoder Sequence to Sequence Model [Електронний ресурс] - Режим доступа: <https://towardsdatascience.com/understanding-encoder-decoder-sequence-to-sequence-model-679e04af4346>
- 21 Jeffrey L. Elman Finding Structure in Time [Текст] // COGNITIVE SCIENCE 14, 179-211с.
- 22 Andrej Karpathy The Unreasonable Effectiveness of Recurrent Neural Networks [Електронний ресурс] - Режим доступа: <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>

ДОДАТОК А

Таблиця А1. СКВ при тренуванні LSTM мережі в залежності від кількості прихованих шарів та кількості ітерацій навчання

Шар\Ітерація	150	155	160	165	170	175	180	185	190	195
150	0.289339	0.368614	0.548057	0.645644	0.257503	0.577637	0.584423	0.437641	0.702882	0.26462
155	0.319121	0.373441	0.221654	0.482442	0.558712	0.312369	0.460781	0.231547	0.260795	0.499809
160	0.238086	0.439906	0.274604	0.267485	0.354373	0.225421	0.32772	0.115237	0.264281	0.138026
165	0.390362	0.461637	0.527306	0.169272	0.539161	0.177739	0.471039	0.23053	0.202315	0.198715
170	0.199491	0.084115	0.346917	0.539044	0.581833	0.233785	0.101128	0.505764	0.110856	0.228848
175	0.131975	0.413144	0.565476	0.247499	0.57332	0.13198	0.379481	0.013801	0.345737	0.440607
180	0.482321	0.496974	0.539943	0.521497	0.371695	0.523984	0.507547	0.686973	0.430542	0.492949
185	0.642928	0.697212	0.489336	0.547644	0.405366	0.447225	0.505847	0.043912	0.239303	0.556729
190	0.545145	0.578841	0.417306	0.179974	0.474376	0.58939	0.285776	0.331728	0.477303	0.399133
195	0.495085	0.433427	0.487317	0.480697	0.535784	0.495865	0.485714	0.458981	0.088221	0.268798
200	0.281118	0.552556	0.481402	0.588713	0.527252	0.474309	0.506909	0.46091	0.511063	0.480954
205	0.429099	0.537107	0.478964	0.548308	0.228108	0.583059	0.358367	0.476865	0.593226	0.421081
210	0.508617	0.492045	0.520127	0.517455	0.577591	0.432259	0.428088	0.542306	0.596629	0.44624
215	0.574689	0.485038	0.152115	0.148551	0.538727	0.664391	0.441695	0.422946	0.465358	0.55181
220	0.254085	0.512796	0.555429	0.299414	0.664683	0.489259	0.277953	0.313998	0.445834	0.413718
225	0.442806	0.350408	0.226759	0.619621	0.367249	0.54342	0.289471	0.458122	0.515918	0.339522
230	0.15182	0.559673	0.292395	0.529972	0.048332	0.462259	0.523672	0.473276	0.467768	0.403885
235	0.373962	0.590259	0.342801	0.26291	0.436661	0.453418	0.407691	0.15352	0.53246	0.177413
240	0.140953	0.632319	0.207036	0.375501	0.532507	0.474636	0.514202	0.546797	0.550973	0.402081
245	0.553456	0.188744	0.122954	0.664737	0.075415	0.559425	0.383011	0.48773	0.040708	0.49863
Шар\Ітерація	200	205	210	215	220	225	230	235	240	245
150	0.61794	0.40312	0.163178	0.31813	0.214089	0.374389	0.1804	0.353661	0.129038	0.305355
155	0.228839	0.470445	0.180906	0.327992	0.373739	0.23565	0.029359	0.164579	0.251051	0.496553
160	0.421985	0.183335	0.156367	0.168747	0.243861	0.564414	0.321513	0.178285	0.375995	0.182826
165	0.234318	0.278483	0.374319	0.50503	0.297911	0.292686	0.483429	0.232484	0.204158	0.227167
170	0.126421	0.350758	0.453097	0.268838	0.261743	0.319074	0.114693	0.049254	0.483437	0.113914
175	0.536407	0.28525	0.341891	0.100375	0.392933	0.384861	0.31468	0.089288	0.044762	0.288166
180	0.339997	0.508966	0.578086	0.453698	0.414761	0.297229	0.508297	0.310192	0.445311	0.174257
185	0.198342	0.474704	0.187703	0.361347	0.580796	0.509924	0.143377	0.493014	0.196803	0.313069
190	0.521177	0.535322	0.351785	0.204594	0.45803	0.441555	0.369715	0.164052	0.363856	0.309905
195	0.225811	0.527943	0.501923	0.38406	0.551882	0.524461	0.509724	0.25485	0.511146	0.485017
200	0.521908	0.264473	0.555136	0.478006	0.159874	0.470694	0.408066	0.535134	0.436005	0.527281
205	0.102307	0.130967	0.674197	0.549013	0.482695	0.44516	0.522797	0.496528	0.552942	0.438642
210	0.179698	0.486189	0.364136	0.412951	0.237493	0.395095	0.496489	0.524966	0.032959	0.465469
215	0.439767	0.461666	0.179708	0.376832	0.498821	0.443285	0.548256	0.549421	0.552219	0.459403
220	0.679696	0.160448	0.504501	0.468058	0.52165	0.367223	0.463243	0.485657	0.365826	0.069222
225	0.462273	0.380705	0.489386	0.458298	0.416404	0.413575	0.356794	0.437888	0.153404	0.414795
230	0.561264	0.145331	0.495391	0.322239	0.626938	0.378917	0.520248	0.354231	0.340395	0.347448
235	0.075148	0.509223	0.379924	0.417821	0.434779	0.289568	0.393086	0.042321	0.050216	0.308999
240	0.424277	0.348987	0.151631	0.432508	0.468625	0.341832	0.472521	0.461284	0.34824	0.301444

245	0.331936	0.446878	0.378141	0.283639	0.457706	0.254092	0.416347	0.048386	0.498544	0.45002
-----	----------	----------	----------	----------	----------	----------	----------	----------	----------	---------

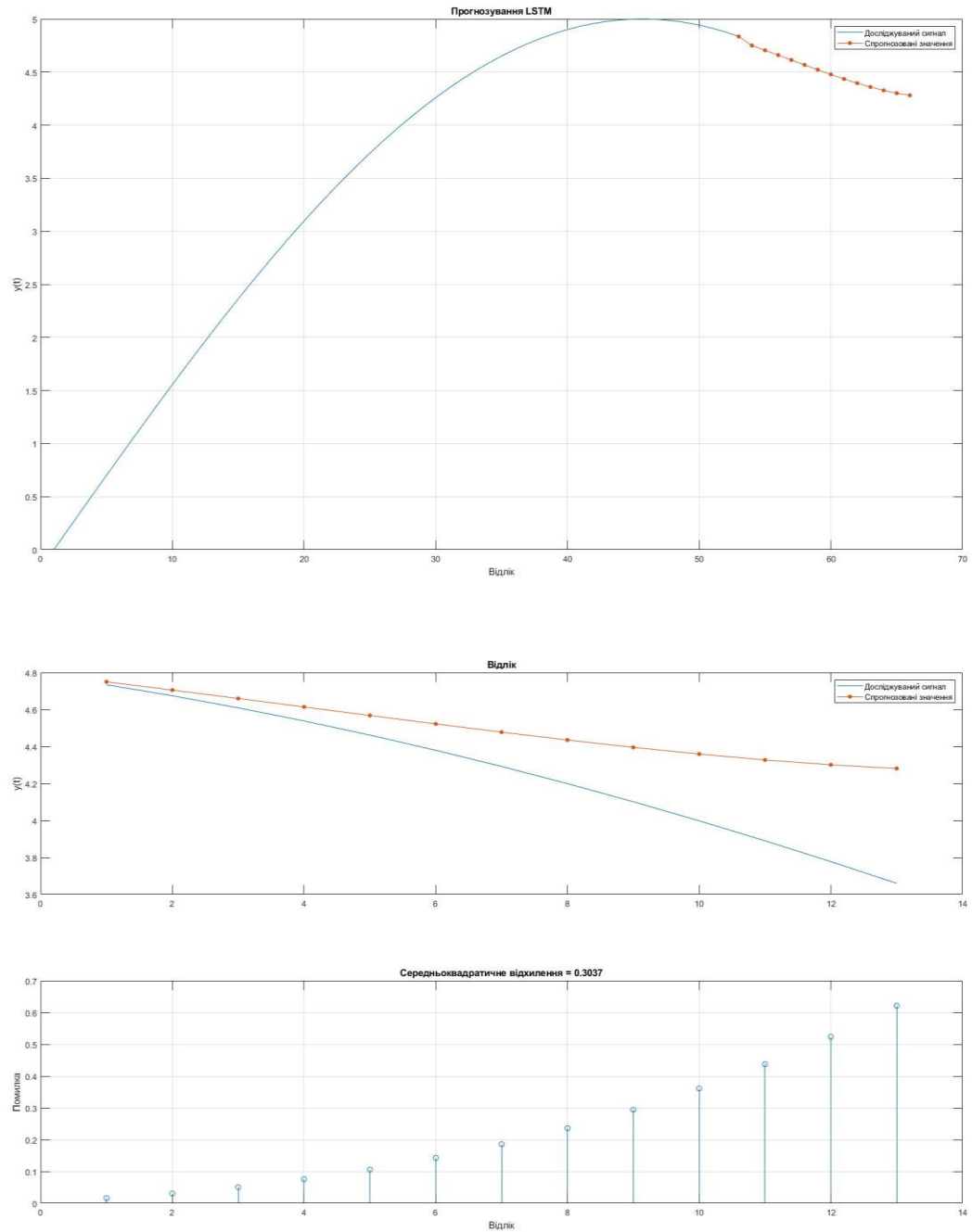


Рис. А1 Результат та помилка роботи LSTM нейронної мережі з 250 прихованих шарів та ітерацій навчання

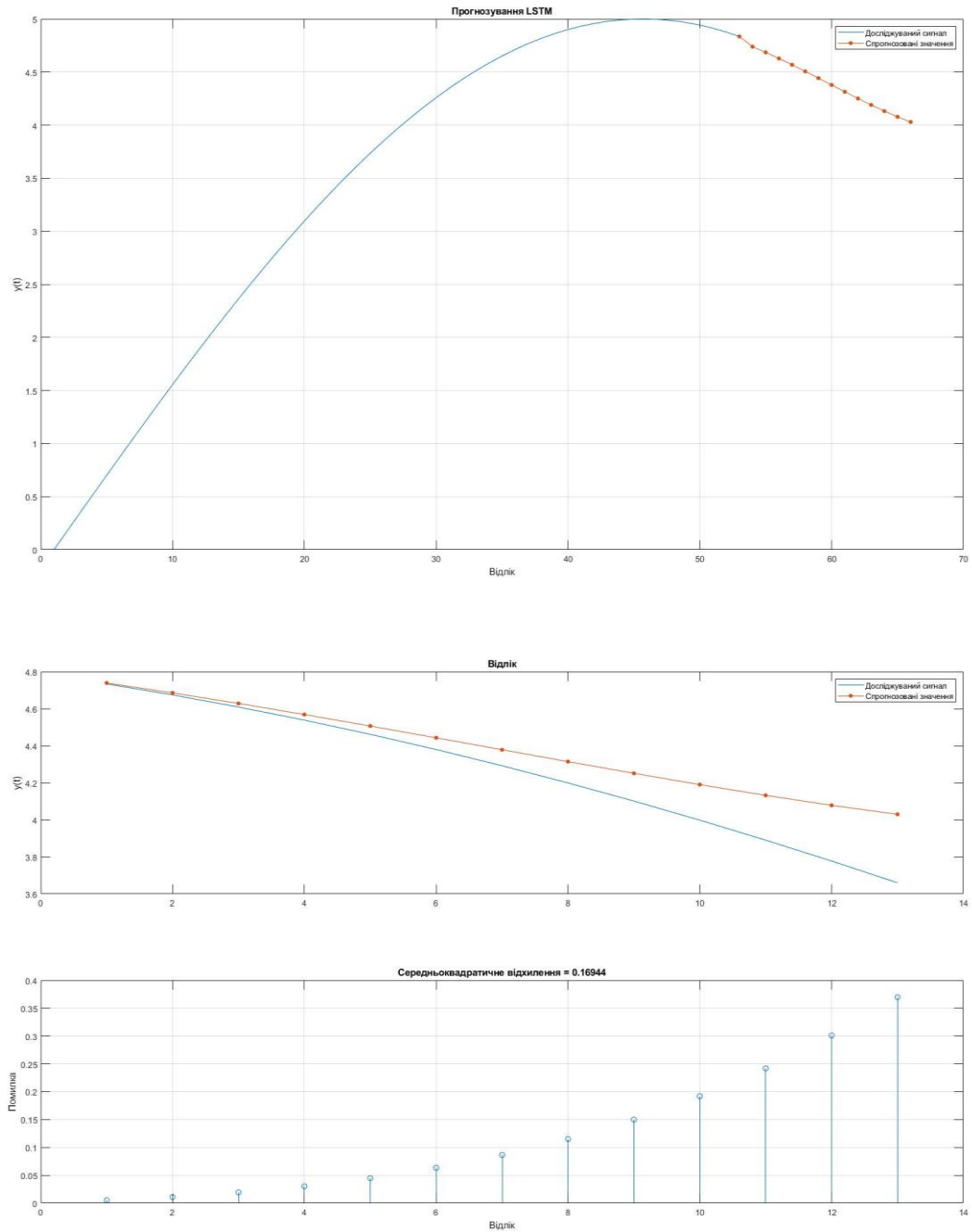


Рис. А2 Результат та помилка роботи LSTM нейронної мережі з 250 прихованих шарів та 200 ітерацій навчання

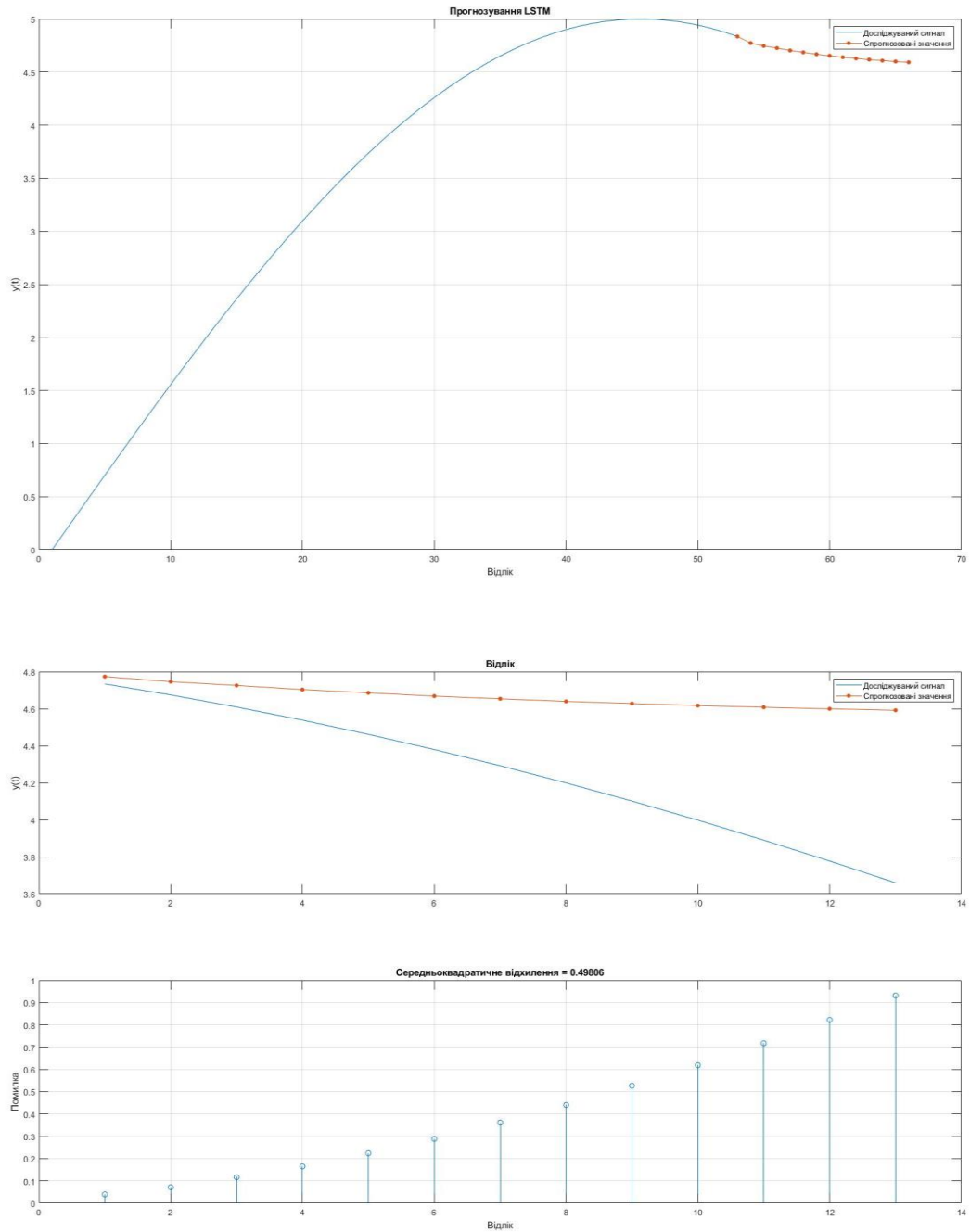


Рис. А3 Результат та помилка роботи LSTM нейронної мережі з 250 прихованих шарів та 150 ітерацій навчання

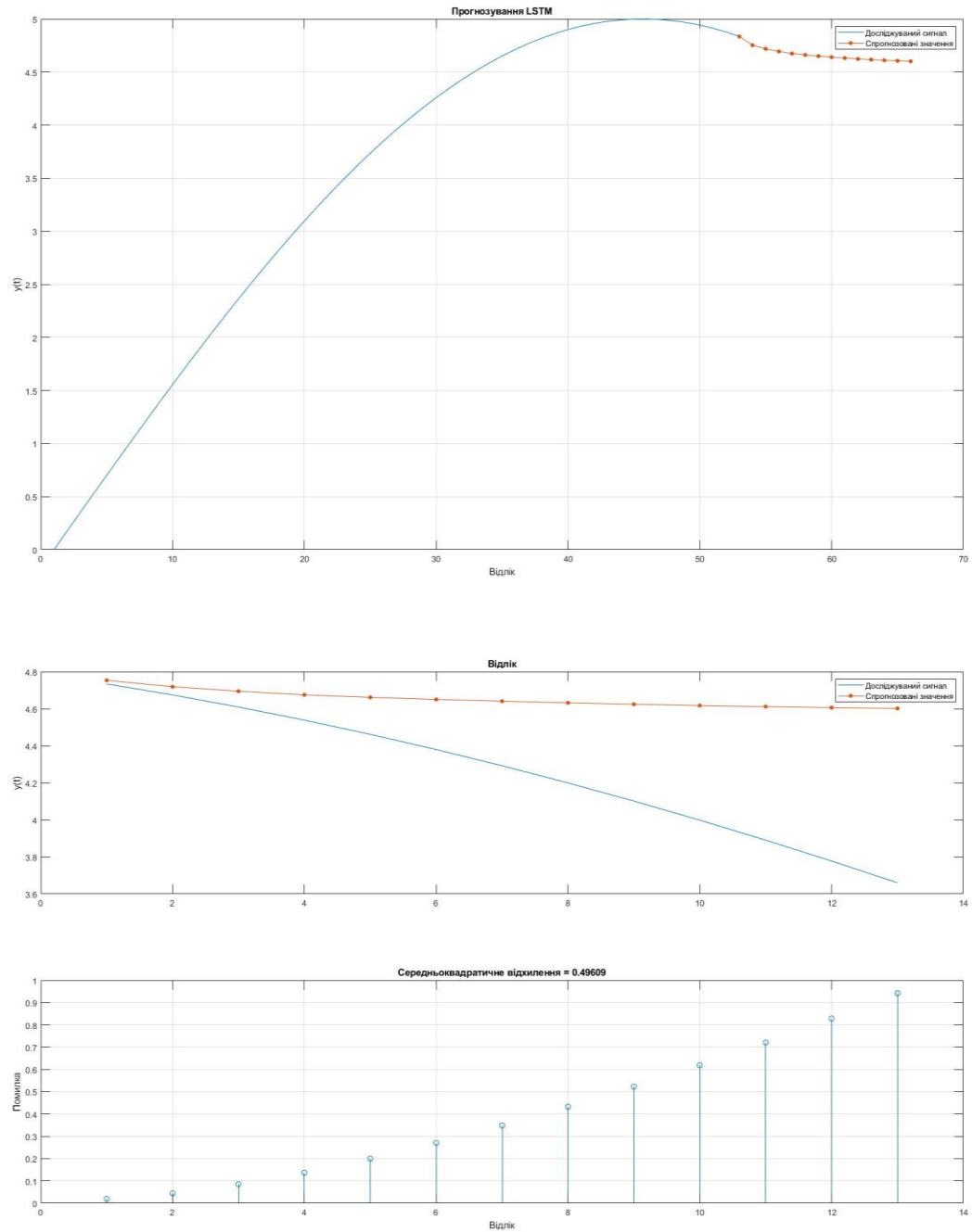


Рис. А4 Результат та помилка роботи LSTM нейронної мережі з 200 прихованих шарів та 200 ітерацій навчання

ДОДАТОК Б

```

clc, clear all

%% Лінійна нейронна мережа

% формування даних
Tk = 5;
w = 0.47;
A = 5;
f = w/(2*pi);
fd = f;
t = 0:fd:Tk;
k = 5;
x = A*sin(w.*t); % - частина синусоїди

a = 2;
% x = 1-exp(-a.*t); % - монотонна експотеніальна функція

% x = k.*t;

P = 1:1:length(t); % - формування відліків часового ряду
T = x;

% 0/3-1/3
% P = P(1:floor(length(t)*1/3));
% T = x(1:floor(length(t)*1/3));
% 1/3-2/3
% P = P(floor(length(t)*1/3):floor(length(t)*2/3));
% T = x(floor(length(t)*1/3):floor(length(t)*2/3));
% 2/3-3/3
% P = P(floor(length(t)*2/3):end);
% T = x(floor(length(t)*2/3):end);

% конфігурація НМ
net = newlind(P,T);
net.IW{1,1}
net.b{1}
%симуляція
%net.layers{1}.transferFcn = 'logsig';
y = sim(net,P);
view(net);

```

```

figure(1)
plot(P,T,'x',P,y,'.-
','LineWidth',0.1,'MarkerSize',15),grid;
title('Робота простої лінійної НМ');
xlabel('Відлік');
ylabel('');
legend('Цільовий сигнал', 'Лінійна регресія НМ');
set(gca,'FontSize',16)

figure(2)
e = T - y;
error = sqrt(mean((T - y).^2)) % - скв

stem(P,e,'x','LineWidth',0.1,'MarkerSize',15),grid;
title('Абсолютна помилка');
xlabel('Відлік');
ylabel('');
set(gca,'FontSize',16)

% % розрахунок функції критерія якості
% w_range = -1:0.1:0.1;
% b_range = 0:0.1:2;
% ES = errsurf(P,T,w_range,b_range,'purelin');
% contour(w_range,b_range,ES,20);
% hold on;
% plot(0.0609,1.5249,'x');
%
% % Будування поверхні функції критерія якості
%
% surfc(w_range,b_range,ES);
clc, clear all

%% LSTM мережа
Tk = 5;
w = 0.47;
A = 5;
f = w/(2*pi);
fd = f;
t = 0:fd:Tk;
x = A*sin(w.*t);
data = x;

numTimeStepsTrain = floor(0.8*numel(data));
dataTrain = data(1:numTimeStepsTrain+1);
dataTest = data(numTimeStepsTrain+1:end);

```

```

mu = mean(dataTrain);
sig = std(dataTrain);

dataTrainStandardized = (dataTrain - mu) / sig;

XTrain = dataTrainStandardized(1:end-1);
YTrain = dataTrainStandardized(2:end);

numFeatures = 1;
numResponses = 1;
numHiddenUnits = 200;

layers = [ ...
    sequenceInputLayer(numFeatures)
    lstmLayer(numHiddenUnits)
    fullyConnectedLayer(numResponses)
    regressionLayer];
options = trainingOptions('adam', ...
    'MaxEpochs',150, ...
    'GradientThreshold',1, ...
    'InitialLearnRate',0.005, ...
    'LearnRateSchedule','piecewise', ...
    'LearnRateDropPeriod',125, ...
    'LearnRateDropFactor',0.2, ...
    'Verbose',0, ...
    'Plots','training-progress');
net = trainNetwork(XTrain,YTrain,layers,options);
dataTestStandardized = (dataTest - mu) / sig;
XTest = dataTestStandardized(1:end-1);
net = predictAndUpdateState(net,XTrain);
[net,YPred] = predictAndUpdateState(net,YTrain(end));

numTimeStepsTest = numel(XTest);
for i = 2:numTimeStepsTest
    [net,YPred(:,i)] =
predictAndUpdateState(net,YPred(:,i-
1),'ExecutionEnvironment','cpu');
end
YPred = sig*YPred + mu;
YTest = dataTest(2:end);
rmse = sqrt(mean((YPred-YTest).^2));

figure
plot(dataTrain(1:end-1))

```

```

hold on
idx =
numTimeStepsTrain:(numTimeStepsTrain+numTimeStepsTest);
plot(idx,[data(numTimeStepsTrain) YPred],'.-
','LineWidth',0.1,'MarkerSize',15),grid;
hold off
xlabel("Відлік")
ylabel("y(t)")
title("Прогнозування LSTM")
legend(["Досліджуваний сигнал" "Спрогнозовані значення"])

figure
subplot(2,1,1)
plot(YTest,'LineWidth',0.1,'MarkerSize',15),grid;
hold on
plot(YPred,'.-','LineWidth',0.1,'MarkerSize',15),grid;
hold off
legend(["Досліджуваний сигнал" "Спрогнозовані значення"])
ylabel("y(t)")
title("Відлік")

subplot(2,1,2)
stem(YPred - YTest),grid;
xlabel("Відлік")
ylabel("Помилка")
title("Середньоквадратичне відхилення = " + rmse)

```